Computational Complexity of One-Step Methods for the Numerical Solution of Initial Value Problems

Carnegie-Mellon Univ Pittsburgh Pa Dept of Computer Science

Sep 76

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A032346

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPUTATIONAL COMPLEXITY OF ONE-STEP METHODS FOR THE NUMERICAL SOLUTION OF INITIAL VALUE PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Arthur Gustav Werschulz		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0370
9. PERFORMING ORGANIZATION NAME AND ADDRESS Carnegie-Mellon University Computer Science Dept. Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE September 1976 13. NUMBER OF PAGES 89
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the task of numerically approximating the solution of an ordinary differential equation initial-value problem. We are interested in two questions:

(1.) For any given one-step method, what is the complexity of finding an approximate solution with error less than ε?

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

(2.) Given an infinite sequence of one-step methods of increasing order, how should the method and the step-size be picked so as to minimize the complexity of finding such an approximation?

We describe a methodology that handles both questions. Furthermore, we find that within such a sequence of methods, the following hold under very general circumstances:

(1.) For any ε, $0 < \varepsilon < 1$, there is a unique choice of order and step-size which minimizes the complexity.

(2.) As ε decreases, both the optimal order and the complexity increase monotonically, tending to infinity as ε tends to zero.

These results are applied to several classes of one-step methods. In doing so, we exhibit some new Taylor series methods that are asymptotically better than Runge-Kutta methods for problems of small dimension. Moreover, we prove that among all classes of nonlinear Runge-Kutta methods, those due to Brent have the highest order possible.

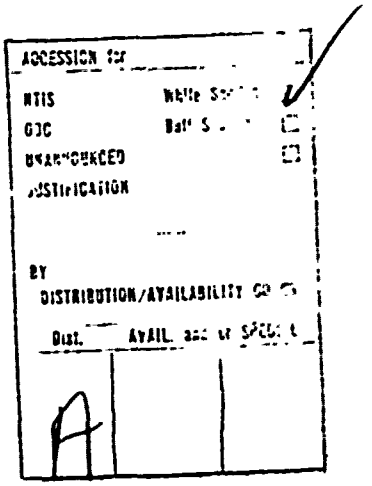
Computational Complexity of One-Step Methods for the Numerical Solution of Initial Value Problems

Arthur Gustav Werschulz

September, 1976

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie-Mellon University

Department of Mathematics
Carnegie-Mellon University
Pittsburgh, PA 15213



This research was supported in part by the National Science Foundation under Grant MCS75-222-55 and the Office of Naval Research under Contract N00014-76-C-0370, NR 044-422.

Acknowledgements

I would like to express my gratitude to those who have helped to make my graduate years at Carnegie-Mellon University just a bit more pleasant. My deepest thanks go to my thesis advisor, Professor Joseph F. Traub, for all the guidance given me and time spent on my behalf.

I am also indebted to Professors George J. Fix, Richard C. MacCamy, and H. T. Kung (Carnegie-Mellon University) and Professor Henryk Woźniakowski (University of Warsaw) for their helpful comments and suggestions as members of my examination committee. In addition, I would like to thank Professor Richard P. Brent (Australian National University) for the interest he has shown in this work.

The Departments of Mathematics and Computer Science have graciously given me financial support during the last four years. In addition, I was spared a certain amount of tedious calculation by the use of the MACSYMA System, developed by the Mathlab group at the Massachusetts Institute of Technology, which is supported by Defense Advanced Research Projects Agency work order 2095, under Office of Naval Research Contract N00014-75-C-0661.

Finally, I would like to thank my wife Patricia for her love and support. Without her, I would have found it difficult to maintain the sense of humor and perspective that is necessary (and perhaps, under certain hypotheses, sufficient) for the successful completion of a doctoral thesis.

Abstract

We consider the task of numerically approximating the solution of an ordinary differential equation initial-value problem. We are interested in two questions:

(1.) For any given one-step method, what is the complexity of finding an approximate solution with error less than ε ?

(2.) Given an infinite sequence of one-step methods of increasing order, how should the method and the step-size be picked so as to minimize the complexity of finding such an approximation?

We describe a methodology that handles both questions. Furthermore, we find that within such a sequence of methods, the following hold under very general circumstances:

(1.) For any ε, $0 < \varepsilon < 1$, there is a unique choice of order and step-size which minimizes the complexity.

(2.) As ε decreases, both the optimal order and the complexity increase monotonically, tending to infinity as ε tends to zero.

These results are applied to several classes of one-step methods. In doing so, we exhibit some new Taylor series methods that are asymptotically better than Runge-Kutta methods for problems of small dimension. Moreover, we prove that among all classes of nonlinear Runge-Kutta methods, those due to Brent have the highest order possible.

Table of Contents

Section 1

Introduction

With few exceptions, past work in analytic computational complexity has focused on the problem of finding a zero of a (nonlinear) transformation of Banach spaces; in most work, this problem is specialized to that of finding a zero of an operator on a finite-dimensional real or complex vector space (and in much of this work, the problem is further specialized to the one-dimensional case). Much has been discovered about the computational aspects of iterative schemes for the solution of such problems, especially in the areas of minimal complexity (e.g., Kung and Traub [73], Traub and Woźniakowski [76]) and maximal order (e.g., Kung and Traub [74], Woźniakowski [75]).

In this paper, we will consider another topic in analytic computational complexity theory, that of finding complexity bounds for the numerical solution of ordinary differential equation initial-value problems on a fixed interval. We will not be interested in questions of the existence and the uniqueness of the solutions to such problems; in fact, we will restrict our discussion of the application of general results to the case where the unique solutions to these problems are analytic functions.

We will limit ourselves here to classes of <u>one-step methods</u> for the numerical solution of these problems; in terms of informational usage, these methods are analogous to iterative zero-finding methods without memory (Traub [64], [72]). Analogous to the one-point iterative methods with memory are the <u>multistep methods</u> for initial-value problems; these methods will be dealt with in a future paper.

Our approach will be to assume that an initial-value problem is given, along with

some error criterion ε, where $0 < \varepsilon < 1$; we then wish to compute an approximate solution with error no greater than ε. Two basic questions concern us:

(1.) For any given method, what is the complexity of solving this problem?

(2.) Given any "basic" sequence of methods with increasing order, which method has minimal complexity?

In Section 2, we describe a methodology that handles both questions for classes consisting of methods whose error functions have a special form. Furthermore, we find that within such a basic sequence of methods, the following hold under very general conditions:

(1.) For any ε, there is a unique choice of order and step size minimizing the complexity.

(2.) As ε decreases, both the optimal order and the complexity increase monotonically, tending to infinity as ε tends to zero.

Furthermore, within many classes of problems and methods, the "penalty" (e.g., the amount the cost curve turns near the optimum) associated with using non-optimal order tends to infinity as ε tends to zero.

These conclusions are an interesting contrast to known results on zero-finding via iterations without memory. The latter results tend to support the "folklore" idea that it is "better" to use a low-order method many times, than to use a high-order method a few times. In the one-point case, optimal order is low, while in the multipoint case, optimal order increases with the problem complexity (but with little penalty for using a method of non-optimal order) (Kung and Traub [73]). In addition, optimal order for these problems does not depend on the error criterion; it is computed for the limiting case as ε approaches zero.

One may wonder why there is this discrepancy between the results for the initial-value problem and those for the zero-finding problem, since any initial-value

problem may be written as an operator equation, as in Stetter [73]. The reason for this is that the methods used for the two problems differ greatly--those for the initial-value problem compute estimates for solution values at new points by discretization, while those for zero-finding compute improved estimates for the zero of a function by iteration.

In Section 3, we discuss the extension of these results to classes consisting of methods whose error functions are somewhat more complicated than those considered in Section 2.

In Section 4, we introduce the notions of normality and order-convergence for a basic sequence of one-step methods. We prove that they are equivalent under certain circumstances. A basic sequence of methods enjoying these properties is very easy to deal with in many respects, especially when one is interested in comparing upper and lower complexity bounds for such a class.

In Section 5, we apply the theory developed in the preceeding sections to the general problem of an autonomous system of equations. We show that the optimal order and complexity behave as described by (1.) and (2.) above for the class of Taylor series methods and for various classes of Runge-Kutta methods. In addition, we construct new Taylor series methods that are asymptotically better (as ϵ tends to zero) than Runge-Kutta methods for problems of small dimension.

In Section 6, we look at the problem of a single scalar autonomous equation. In this case, we may use the classes of "nonlinear Runge-Kutta methods" developed by Brent [74], [76]. We show that the behavior described by (1.) and (2.) above holds for these methods. In addition, we prove that among all classes of nonlinear Runge-Kutta methods, those due to Brent have the highest order possible.

Section 7 describes some numerical data that support the above theoretical results. In particular, these data seem to indicate that even for modest values of s, there are considerable savings in using methods of optimal, rather than fixed, order.

Finally, in Section 8, we draw some conclusions, make some comparisons, point out some unanswered questions, and define new areas to which this theory should be extended.

Section 2

Optimality Within a Strong Basic Sequence

We are interested in the numerical solution of a class of ordinary differential equation initial-value problems on a fixed interval I of finite length; we take $I = [0, 1]$ without loss of generality. More precisely, let $\mathcal{D}$ be a set of initial-value points in the real N-dimensional linear space $\mathbb{R}^N$, and let $\mathcal{V}$ be a set of operators on $\mathbb{R}^N$, such that the initial-value problem of finding a function $x: I \to \mathbb{R}^N$ satisfying

$$\begin{aligned} \dot{x}(t) &= v(x(t)) \quad \text{if } t \in \text{int } I, \\ x(0) &= x_0 \end{aligned} \tag{2.1}$$

has a unique solution for every $(x_0, v) \in \mathcal{D}\times\mathcal{V}$. The autonomous form of this system is no restriction, since any non-autonomous system may be made autonomous by increasing the dimension of the system by one.

The model of computation to be used is fairly general. We assume only that all arithmetic operations are performed exactly in $\mathbb{R}$ (i.e., infinite-precision arithmetic), and that for any algorithm to be considered for the solution of (2.1), a set of procedures is given for the computation of any information about v required by that algorithm. (For instance, with Runge-Kutta methods, we must be able to compute v at any point in its domain.)

In this paper, we are interested in the numerical solution of (2.1) via <u>one-step methods</u>, using an <u>equidistant grid</u> as defined in Stetter [73]. (We limit ourselves to equidistant grids in order to facilitate the comparison of methods of different orders; the other extreme is taken by Lindberg [74], who considers the problem of picking an

optimal grid for a given method of fixed order.) Thus the methods considered will generate approximations x_i to $x(t_i)$ by the recursion

$$x_{i+1} = x_i + h\,\varphi(x_i, h) \qquad (0 \le i \le n-1,\ n = h^{-1}) \tag{2.2}$$

where h is the step-size and φ is the increment function for the method (Henrici [62]); for briefness, we will refer to "the method φ." Despite the fact that $\varphi(x_i, h)$ will depend on some information about v, we will not explicitly indicate this dependence.

Thus, the method φ produces an approximation to the true solution of (2.1). We want to measure the discrepancy between the approximate and true solutions. Various error measures have been introduced in the literature. These include the local truncation error per step, the local truncation error per unit step, and the global error; see Henrici [62] or Stetter [73] for definitions. These error measures may be either absolute or relative (in the usual sense); they may be measured either at the endpoint of the interval (as in Henrici [62], Hindmarsh [74]) or over the entire grid (as in Sandberg [67], Lindberg [74]). There has been a great deal of discussion of which error criterion is the best one to use; for instance, Gear [71] (Section 9.3) uses local error per step, while Hull et al. [72] use local error per unit step. We take no sides in this discussion, since any of these error measures may be used in the analysis to follow.

Before proceeding any further, we will establish some notational conventions. Let $\mathfrak{X}$ be an ordered ring; then $\mathfrak{X}^+$ and $\mathfrak{X}^{++}$ will respectively denote the nonnegative and positive elements of $\mathfrak{X}$. (This will be used in the cases $\mathfrak{X} = \mathbb{R}$, the real numbers, and $\mathfrak{X} = \mathbb{Z}$, the integers.) The symbol ":=" means "is defined to be," while "$\equiv$" means "is identically equal to." The symbol "∇" will be used to denote the gradient of a mapping. If $\chi_1, \chi_2: \mathbb{R} \to \mathbb{R}$ and $\omega: \mathbb{R}^2 \to \mathbb{R}$ are differentiable, then for $i = 1, 2$, we will write

$$\partial_i \omega(x_1(t),x_2(t))$$

for the result of differentiating $\omega(x_1,x_2)$ with respect to x_i, and then substituting $x_1 = x_1(t)$, $x_2 = x_2(t)$. We use the notations "$x \downarrow a$" and "$x \uparrow a$" to indicate one-sided limits as in Burr [65]. Finally, we shall write "$(a.b)_c$" to inidicate the c^{th} part of equation (a.b), as in Gurtin [75].

Now we are prepared to define our problem. Let $\mathcal{D}$ and $\mathcal{V}$ be as above; consider a problem (x_0,v) in $\mathcal{D}\times\mathcal{V}$. Let Φ be a class of one-step methods, and let $\sigma: \Phi\times I \to \mathbb{R}^+$ satisfying $\lim_{h\downarrow 0} \sigma(\varphi,h) = 0$ be a given function that will serve as an error measure. Choose an error criterion ε satisfying the technical restriction $0 < \varepsilon < 1$. We then wish to answer two questions:

(1.) Given $\varphi \in \Phi$, how may we pick $h \in I$ such that

$$\sigma(\varphi,h) \le \varepsilon, \tag{2.3}$$

and what is the complexity of the process defined by φ and h?

(2.) How may one choose among all $(\varphi,h) \in \Phi\times I$ such that (2.3) holds, that pair (φ^*,h^*) giving minimal complexity?

In order to get useful bounds on $\sigma(\varphi,h)$, it is necessary to introduce the concept of order. In this section, we will use a highly restricted definition, which we will relax in Section 3. Let $\Phi = \{\varphi_p: p \in \mathbb{Z}^{++}\}$, and suppose that there is an analytic function $\kappa: \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{p\to 0} \kappa(p)^{1/p}$ exists and is nonzero and

$$\sigma(\varphi_p,h) = \kappa(p)\, h^p \quad \text{for } h \in I \text{ and } p \in \mathbb{Z}^{++} . \tag{2.4}$$

Then φ_p is said to have strong order p with respect to σ, and Φ is said to be a strong basic sequence. (Although the error coefficient κ will generally depend on the solution x of (2.1), we do not explicitly indicate this dependence.) Note that the order of a method depends on the error measure; for example, the order with respect to the local error per step is one greater than that with respect to the local error per unit step or the global error.

Equation (2.4) is somewhat more restrictive than that which is usually encountered in practice; more often, we expect κ to depend on h. We consider the extension of our results to this case in the next section.

We now are able to measure the complexity of computing an approximate solution to (2.1), with error not exceeding ε, using a strong basic sequence Φ. Indeed, (2.4) implies that a necessary and sufficient condition for $\varphi(\varphi_p,h) = \varepsilon$ is that

$$h = h(p,\alpha) := \kappa(p)^{-1/p}\, e^{-\alpha/p}, \tag{2.5}$$

where

$$\alpha := \ln(\varepsilon^{-1}). \tag{2.6}$$

(Note that since $0 < \varepsilon < 1$, we have $\alpha \in \mathbb{R}^{++}$.) Thus, the number of steps needed is given by

$$n := h^{-1} = \kappa(p)^{1/p}\, e^{\alpha/p}. \tag{2.7}$$

(Note that n (as given by (2.7)) need not be an integer. But this poses no essential difficulty; see (e.g.) Traub and Woźniakowski [76].) Next, suppose that there exists an analytic function $c: \mathbb{R}^+ \to \mathbb{R}^+$ such that c(p) is the cost per step associated with the method φ_p. Finally, we assume that the cost per step does not vary from step to step; for the classes of methods we consider, this means only that we assume that the cost of evaluating v (or its derivatives) does not depend on the point of evaluation. Thus the complexity $C(p,\alpha)$ of solving (2.1) to within an error criterion $\varepsilon = e^{-\alpha}$ is simply given by

$$C(p,\alpha) = n\,c(p) = f(p)\, e^{\alpha/p}, \tag{2.8}$$

where we define $f: \mathbb{R}^+ \to \mathbb{R}^+$ by

$$f(p) := \kappa(p)^{1/p}\, c(p). \tag{2.9}$$

We now turn to the question of picking for each $\alpha \in \mathbb{R}^{++}$ that order p giving

minimal complexity. In the analysis to follow, we will drop the restriction that p must be an integer. However, we will recover optimality over the integers from optimality over the real numbers in Corollary 2.1 . Without loss of generality, we assume that

(2.10) $$p > 0 \text{ implies } f(p) > 0 .$$

(If there were a $p > 0$ with $f(p) = 0$, use of the method φ_p would yield a solution with zero complexity, i.e., "with no effort.") In addition, we assume that

(2.11) $$\lim_{p\uparrow\infty} f(p) = +\infty .$$

By (2.9), this assumption maybe viewed as a simple consequence of two conditions, both of which are quite natural. The first is that $\lim_{p\uparrow\infty} c(p) = +\infty$; the "better" a method is (i.e., the higher its order is), the more we should expect to pay for its use. The second condition is that if $\lim_{p\uparrow\infty} \kappa(p) = 0$, then there must exist a $\beta \in I$ such that $\kappa(p) \geq \beta^p$ for p sufficiently large. (For example, in the class of Taylor series methods, using the worst-case local error per unit step as the error measure, this second condition would follow from the assumption that any problem $(x_0,v) \in \mathcal{D}\times\mathcal{V}$ must have an analytic solution.)

Thus in order to find a minimum for $C(\cdot,\alpha)$, we merely differentiate (2.8) with respect to p, finding

(2.12) $$\partial_1 C(p,\alpha) = p^{-2} f(p) e^{\alpha/p} [G(p) - \alpha] ,$$

where $G: \mathbb{R}^{++} \to \mathbb{R}$ is given by

(2.13) $$G(p) := p^2 f'(p)/f(p) .$$

Thus a necessary condition that p be a minimum for $C(\cdot,\alpha)$ is that $\partial_1 C(p,\alpha) = 0$, i.e.,

(2.14) $$G(p) = \alpha .$$

Sufficient conditions for the existence and uniqueness of a p satisfying (2.14) and minimizing $C(\cdot,\alpha)$ are given in

Theorem 2.1: Let f satisfy (2.10) and (2.11). Suppose also that

(2.15) $\qquad G'(p) > 0 \text{ whenever } G(p) > 0.$

Then there is a function $p^*: \mathbb{R}^{++} \to \mathbb{R}^{++}$ such that (2.14) holds if and only if $p = p^*(\alpha)$. Moreover, for all $p \in \mathbb{R}^{++}$,

(2.16) $\qquad C^*(\alpha) := C(p^*(\alpha),\alpha) \le C(p,\alpha),$

with equality holding if and only if $p = p^*(\alpha)$.

(Since $p^*(\alpha)$ satisfies (2.15), we call $p^*(\alpha)$ the optimal order, $C^*(\alpha)$ the optimal complexity, and

(2.17) $\qquad h^*(\alpha) := h(p^*(\alpha),\alpha)$

the optimal step-size.)

Proof of Theorem 2.1: If we write the Maclaurin series of f and substitute it into (2.13), it is easy to see that

(2.18) $\qquad \lim_{p \downarrow 0} G(p) = 0.$

We now claim that

(2.19) $\qquad \lim_{p \uparrow \infty} G(p) = +\infty.$

Indeed, since (2.11) holds there is a $p_0 > 0$ such that $f'(p_0) > 0$, i.e., $G(p_0) > 0$. Thus by (2.15), G is monotone increasing on $[p_0, +\infty)$, and hence either (2.19) holds or there exists a $\gamma > 0$ such that $\lim_{p \uparrow \infty} G(p) = \gamma$. If the latter holds, then G is bounded, and we have

$$f'(t)/f(t) \le \delta t^{-2} \qquad (1 \le t < +\infty)$$

for some $\delta > 0$; integrating the above inequality over $1 \le t \le p$ yields

$$f(p) \le f(1)\, e^{\delta(1 - 1/p)},$$

so that $\lim_{p \uparrow \infty} f(p) \le f(1)\, e^{\delta}$, contradicting (2.11). Thus (2.18) and (2.19) hold; together, they imply that for any $\alpha > 0$, there is a choice of p such that (2.14) holds.

Suppose that for some $\alpha > 0$, there were two numbers $p_0 < p_1$ with $G(p_0) = G(p_1) = \alpha$. Then by Rolle's Theorem, there is a p_2 between p_0 and p_1 with $G'(p_2) = 0$, contradicting (2.15). Thus for each $\alpha > 0$, there is a unique choice of p such that (2.14) holds; we denote this choice by $p^*(\alpha)$.

To prove (2.16), differentiate (2.12) with respect to p to find

$$\partial_1^2 C(p,\alpha) = p^{-2} f(p) e^{\alpha/p} G'(p) + [G(p) - \alpha] (\partial/\partial p) [p^{-2} f(p) e^{\alpha/p}] . \tag{2.20}$$

But upon substituting $p = p^*(\alpha)$, the second term in (2.20) vanishes and the first term is positive; so we have

$$\partial_1^2 C(p^*(\alpha),\alpha) > 0 .$$

Thus $p^*(\alpha)$ gives a local minimum for $C(\cdot,\alpha)$, which has only one critical point (since (2.14) has a unique solution) and (2.16) follows. ∎

Note that we have not said that $p^*(\alpha)$ is an integer; in fact, this need not be true in general. Since the basic sequence Φ is indexed by $\mathbb{Z}^{++}$, we have not yet solved the problem of choosing from among all (φ_p,h) such that (2.3) holds, that pair yielding minimal complexity. This problem is solved by

Corollary 2.1: For any $\alpha > 0$, define $p^{**}(\alpha) \in \mathbb{Z}^{++}$ to be that element of the set $\{\lfloor p^*(\alpha) \rfloor , \lceil p^*(\alpha) \rceil\}$ which gives the smaller value of $C(\cdot,\alpha)$. Then

$$C(p^{**}(\alpha),\alpha) \le C(p,\alpha) \quad \text{for } p \in \mathbb{Z}^{++} ,$$

with equality if and only if $p = p^{**}(\alpha)$.

Proof: Clearly we need only consider the case where $p^*(\alpha)$ is not an integer. Suppose there exists $p_0 \in \mathbb{Z}^{++}$, not equal to $p^{**}(\alpha)$, with $C(p_0,\alpha) \le C(p^{**}(\alpha),\alpha)$. Without loss of generality, assume $p_0 < \lfloor p^*(\alpha) \rfloor$. Then $C(p_0,\alpha) \le C(\lfloor p^*(\alpha) \rfloor,\alpha) \ge C(p^*(\alpha),\alpha)$, which implies that there is a $p_1 \in (p_0 , p^*(\alpha))$ such that $\partial_1 C(p_1,\alpha) = 0$. Hence, $G(p_1) = \alpha$, but $p_1 \neq p^*(\alpha)$. This contradicts Theorem 2.1. ∎

It may be readily verified that the hypotheses of Theorem 2.1 are satisfied for many classes of functions f. Some of these are

logarithmic: $f(p) = \ln (p + e)$,

monomial: $f(p) = p^m$ $(m \in \mathbb{R}^{++})$,

exponential: $f(p) = \beta^p$ $(\beta > 1)$,

super-exponential: $f(p) = p^p$, and

hyper-exponential: $f(p) = \beta^{p^p}$.

(We write "ln (p + e)", where e is the base of the natural logarithms, rather than "ln p" as a technical convenience. However, an expression of the form "ln (p + γ)" with $\gamma > 0$ is necessary to guarantee that $f(1) > 0$.) Furthermore, we find that if f has the monomial-logarithmic form

$$f(p) = p^a (\ln (p + e))^b \qquad (a, b \in \mathbb{R}^{++}),$$

then the hypotheses of Theorem 2.1 hold. This may be verified either directly, or by using the following Lemma, along with the fact that the hypotheses hold for $f(p) = p$ and $f(p) = \ln (p + e)$.

Lemma 2.1: Let f have the form

$$f(p) = a \prod_{i=1}^{m} (f_i(p))^{r_i},$$

where $a \in \mathbb{R}^{++}$, and for each i $(1 \le i \le m)$, f_i satisfies the hypotheses of Theorem 2.1 and $r_i \in \mathbb{R}^{++}$. Then f satisfies the hypotheses of Theorem 2.1.

Proof: It is clear that if each f_i satisfies (2.10) and (2.11), then so does f. If each f_i yields (via (2.13)) a G_i satisfying (2.15), then f yields a G in the form

$$G(p) = \sum_{i=1}^{m} r_i G_i(p),$$

and so it is clear that G satisfies (2.15). ∎

Note that for our purposes, we will only be interested in monomial and

monomial-logarithmic growth. We include the other examples of functions that satisfy the hypotheses of Theorem 2.1 to illustrate the wide variety of functions that qualify.

So we have seen that under the hypotheses of Theorem 2.1, there is a unique choice of order and step size minimizing the total complexity for any error criterion. What happens to these choices as α changes?

Theorem 2.2: Let f satisfy the hypotheses of Theorem 2.1. Then

(1.) $p^*(\alpha)$ and $C^*(\alpha)$ increase monotonically with α.

(2.) $\lim_{\alpha\uparrow\infty} p^*(\alpha) = \lim_{\alpha\uparrow\infty} C^*(\alpha) = +\infty$.

(3.) If there exists $M > 0$ such that $\kappa(p)^{1/p} \leq M$ for all p, then $\liminf_{\alpha\uparrow\infty} h^*(\alpha) > 0$ if $\alpha/p^*(\alpha)$ is bounded as $\alpha\uparrow\infty$.

Proof: To prove (1.), note that p^* is the functional inverse of G. Thus $p^{*\prime}(\alpha) = G'(p^*(\alpha))^{-1} > 0$, so that $p^*(\alpha)$ increases with α. Now use the chain rule:

$$C^{*\prime}(\alpha) = \partial_1 C(p^*(\alpha),\alpha)\, p^{*\prime}(\alpha) + \partial_2 C(p^*(\alpha),\alpha).$$

But the first term on the right-hand side vanishes by the definition of $p^*(\alpha)$. So

$$C^{*\prime}(\alpha) = \partial_2 C(p^*(\alpha),\alpha) = (p^*(\alpha))^{-1} f(p^*(\alpha))\, e^{\alpha/p^*(\alpha)} > 0$$

and $C^*(\alpha)$ increases with α.

Suppose that $\lim_{\alpha\uparrow\infty} p^*(\alpha) \neq +\infty$. Since $p^*(\alpha)$ increases monotonically with α, there is an $L > 0$ such that $\lim_{\alpha\uparrow\infty} p^*(\alpha) = L$. So (2.14) implies that

$$G(L) = \lim_{\alpha\uparrow\infty} G(p^*(\alpha)) = \lim_{\alpha\uparrow\infty} \alpha = +\infty,$$

contradicting the continuity of G. This proves the first part of (2.). Now for any $\alpha > 0$, we have

$$C^*(\alpha) = f(p^*(\alpha))\, e^{\alpha/p^*(\alpha)} > f(p^*(\alpha)).$$

Let $\alpha\uparrow\infty$; then (2.11) and the first part of (2.) imply that the second part of (2.) holds.

To prove (3.), let such an $M > 0$ exist, so that $[h^*(\alpha)]^{-1} \leq M\, e^{\alpha/p^*(\alpha)}$. Then we see that $\liminf_{\alpha\uparrow\infty} h^*(\alpha) > 0$ if $[h^*(\alpha)]^{-1}$ is bounded as $\alpha\uparrow\infty$, which itself is true if $\alpha/p^*(\alpha)$ is bounded as $\alpha\uparrow\infty$. ∎

Therefore under a very general set of conditions, we see that the more accuracy we want in our computed solution, the greater its complexity becomes. Of course, this is just what we would expect. What is somewhat surprising is that the minimal complexity is obtained by letting the order p increase as the error ε decreases, with p increasing without bound as ε tends to zero. Moreover, the last part of the theorem says that not only should the order be increased when trying to obtain a more accurate solution, but that it may actually turn out that the step-size should not be allowed to tend to zero.

We now determine whether we are saving a great deal by using the optimal-order method. This may be thought of in several ways; we will consider how sharply the cost curve turns at the optimum, the cost-difference between using a method of fixed order and a method of optimal order, and the cost-ratio of a fixed-order method to an optimal-order method. We will show that under certain reasonable conditions, all of these measures tend to infinity with α.

How sharply the cost curve turns at the maximum is measured by $\partial_1^2 C(p^*(\alpha),\alpha)$. If we consider five of the growth models mentioned above (e.g., monomial, monomial-logarithmic, exponential, hyper-exponential, and super-exponential), we find that $\partial_1^2 C(p^*(\alpha),\alpha)$ is monotone increasing for α sufficiently large, and tends to infinity with α, with but one exception; in the case of "linear growth" ($f(p) = p$), we find that $\partial_1^2 C(p^*(\alpha),\alpha) \equiv e$. However, in the classes of algorithms we study, the case $f(p) = p$ does not arise, provided that we include "combinatory cost" (see Section 5) in our complexity measure. Thus in general, we find that the "pointedness" of the cost curve near the minimum increases without bound as $\alpha \uparrow \infty$.

Next, we will show that for <u>any</u> f satisfying the hypotheses of Theorem 2.1, the

difference in complexity between using a method of fixed order and a method of optimal order tends to infinity with α.

Proposition 2.2: For any fixed $p_0 \in \mathbb{R}^{++}$ such that $G'(p_0) \geq 0$,

$$\lim_{\alpha \uparrow \infty} [C(p_0,\alpha) - C^*(\alpha)] = +\infty .$$

Proof: Pick α so large that $p^*(\alpha) > p_0$, and let $p_0 < p < p^*(\alpha)$. If we write out the partial derivative in the last term of (2.20), we find that

$$\partial_1{}^2 C(p,\alpha) = p^{-2} f(p) e^{\alpha/p} G'(p) + p^{-4} [\alpha - G(p)] [(\alpha + 2p) - G(p)] f(p) .$$

Since $p_0 < p < p^*(\alpha)$, we have $G(p) < \alpha$; it then follows that $\partial_1{}^2 C(p,\alpha)$ is positive and bounded away from zero as α tends to infinity. Since

$$C(p_0,\alpha) - C^*(\alpha) = \partial_1{}^2 C(p,\alpha) [p_0 - p^*(\alpha)]^2 / 2$$

for some p between p_0 and $p^*(\alpha)$, the result follows. ∎

As for the cost-ratio, a simple calculation shows that

$$\lim_{\alpha \uparrow \infty} C(p_0,\alpha)/C^*(\alpha) = +\infty$$

in all of the examples given above. Thus there are a number of ways in which we incur a large additional cost by not using the optimal order.

One may wonder whether the result that optimal order increases and tends to infinity with α is "reasonable." One way of determining this is to examine actual numerical tests; we cite Hull et al. [72] as a well-known example. Since we are only dealing with methods of fixed order, our theory does not attempt to handle methods such as Bulirsch-Stoer, Krogh, or Gear. However, let us look at the results of Hull et al. for the Runge-Kutta methods (which <u>are</u> germane to our discussion--See Section 5.2). Even though there are only three methods (of orders four, six, and eight) and three error criteria ($\varepsilon = 10^{-3}$, 10^{-6}, and 10^{-9}), Table 1 in Hull et al. [72] indicates that the optimal order <u>does</u> increase as ε decreases. (We give more extensive numerical data in Section 7.)

Finally, we note that the restriction that the grid be equidistant may be weakened somewhat, provided that we use a local error measure. Indeed, let I be partitioned as $I = I_1 \cup \ldots \cup I_L$, and now assume that we use a grid that is equidistant on each subinterval $I_1, \ldots, I_L$. Then the total complexity is given by the sum of the complexities of all subintervals

$$C(p_1,\ldots,p_L,\alpha) := \sum_{i=1}^{L} C_i(p_i,\alpha),$$

where we set

$$C_i(p,\alpha) := f_i(p,\alpha)\, e^{\alpha/p} \quad , \qquad f_i(p) := \kappa_i(p)^{1/p}\, c(p) ;$$

here $\kappa_i(p)$ is the error constant of φ_p on I_i. Since we use a local error measure, we find that $C(p_1,\ldots,p_L,\alpha)$ is minimized by choosing each p_i to minimize $C_i(\cdot\,,\alpha)$. Thus the earlier results apply; in particular, if we define $p_i^*(\alpha)$ to be the optimal order on I_i, we find that if f_i satisfies (2.10), (2.11), and (2.15), then $p_i^*(\alpha)$ increases and tends to infinity with α.

Section 3

Optimality Within a Basic Sequence

There are two difficulties with the approach taken in Section 2. The first has already been mentioned--we generally expect the error coefficient to depend on the step-size. The second is based on the fact that there are a large number of p^{th}-order methods of a given type, and we wish to use the best method possible. In theory, this would involve finding a p^{th}-order method with minimal cost per step. In practice, this is not often possible; there is a gap between the minimal cost theoretically possible and the cost of the best method known. So we now consider the extension of the results in Section 2 to a more general setting, which will take these two difficulties into account.

We first refine our notion of order. Let σ: $\Phi \times I \to \mathbb{R}^+$ be an error measure, where $\Phi = \{\varphi_p : p \in \mathbb{Z}^{++}\}$ is a class of one-step methods, and suppose that a function r: $\mathbb{R}^+ \times I \to \mathbb{R}^+$ and analytic functions $\kappa_L , \kappa_U : \mathbb{R}^+ \to \mathbb{R}^+$ exist such that $\lim_{p\to 0} \kappa_L(p)^{1/p}$ and $\lim_{p\to 0} \kappa_U(p)^{1/p}$ exist and are nonzero and

$$\sigma(\varphi_p,h) = \kappa(p,h)\, h^p \quad \text{for } h \in I \text{ and } p \in \mathbb{Z}^{++}, \tag{3.1}$$

where

$$0 < \kappa_L(p) \le \kappa(p,h) \le \kappa_U(p) < +\infty \quad \text{for } h \in I . \tag{3.2}$$

Then φ_p is said to have <u>order</u> p with respect to σ, and Φ is said to be a <u>basic sequence</u> (as in Traub [64]); $\kappa(p,h)$ is said to be the <u>error coefficient</u> of φ_p. (Here we introduce the convention of attaching the subscripts "L" and "U" to quantities that refer to lower and upper bounds on complexity, respectively.)

This definition of order is similar to that in Cooper [69] and Cooper and Verner [72], except that we include a lower bound $\kappa_L(p)$ on $\kappa(p,h)$; this lower bound is necessary and sufficient to guarantee that the order of a method is well-defined. Note that this definition makes sense for all values of $h \in I$; thus, it is non-asymptotic in that we do not require $h \downarrow 0$ in order for it to make sense. Clearly, a strong basic sequence is a basic sequence; hence, the definition of order is an extension of the definition of strong order given in Section 2. Finally, note that the order depends on the choice of the error measure σ; for instance, the order with respect to the local error per step exceeds that with respect to the local error per unit step by one.

We next discuss the notion of cost per step. As pointed out above, we will generally have only bounds on the cost $c(p)$ required per step of a given p^{th}-order method:

$$c_L(p) \le c(p) \le c_U(p). \tag{3.3}$$

That is, $c_L(p)$ is a lower bound on the minimum possible cost per step, usually derived via theoretical considerations, and $c_U(p)$ is an upper bound on the minimum possible cost per step, which is derived by exhibiting an algorithm for computing φ_p. (In what follows, we shall assume that $c_L, c_U : \mathbb{R}^+ \to \mathbb{R}^+$ are analytic functions.)

We now wish to give bounds on $C(p,\alpha)$, the complexity of finding an approximate solution of (2.1) using the method φ_p, such that $\sigma(\varphi_p,h) \le e^{-\alpha}$. Suppose that (2.3) holds. Then by (3.1) and (3.2), we must have

$$\kappa_L(p)h^p \le e^{-\alpha}, \quad \text{i.e.,} \quad h \le h_L(p,\alpha) := \kappa_L(p)^{-1/p} e^{-\alpha/p}. \tag{3.4}$$

Hence, the number of steps $n = h^{-1}$ must satisfy

$$n \ge \kappa_L(p)^{1/p} e^{\alpha/p}. \tag{3.5}$$

Defining (as in Section 2)

$$C(p,\alpha) := n\, c(p) \tag{3.6}$$

(i.e., total complexity equals number of steps required multiplied by cost required per step), (3.3) and (3.5) imply that

$$C(p,\alpha) \geq C_L(p,\alpha) := f_L(p)\, e^{\alpha/p}, \tag{3.7}$$

where

$$f_L(p) := \kappa_L(p)^{1/p}\, c_L(p). \tag{3.8}$$

That is, regardless of the algorithm used to compute φ_p, the total complexity of finding an approximate solution of (2.1) must exceed $C_L(p,\alpha)$.

On the other hand, we find that in order to use φ_p to find such an approximate solution, it suffices (by (3.1) and (3.2)) to take

$$\kappa_U(p)\, h^p = e^{-\alpha}, \text{ i.e., } h = h_U(p,\alpha) := \kappa_U(p)^{-1/p}\, e^{-\alpha/p}, \tag{3.9}$$

so that we need only take n steps, where

$$n = \kappa_U(p)^{1/p}\, e^{\alpha/p}. \tag{3.10}$$

(As in Section 2, the value of n given by (3.10) need not be an integer; again, this is handled as in Traub and Woźniakowski [76].) Thus (3.3), (3.6), and (3.10) imply that

$$C(p,\alpha) \leq C_U(p,\alpha) := f_U(p)\, e^{\alpha/p}, \tag{3.11}$$

where

$$f_U(p) := \kappa_U(p)^{1/p}\, c_U(p). \tag{3.12}$$

That is, there exists an algorithm for computing φ_p such that the total complexity of finding an approximate solution of (2.1) equals $C_U(p,\alpha)$. We summarize the above results in

Theorem 3.1: Let $C(p,\alpha)$ be the complexity of finding an approximate solution of (2.1), using the method φ_p, with $\sigma(\varphi_p,h) \le e^{-\alpha}$. Then

(3.13) $$C_L(p,\alpha) \le C(p,\alpha) \le C_U(p,\alpha),$$

where C_L and C_U are given by (3.7) and (3.11). Moreover, if $h = h(p,\alpha)$ is the maximal step-size for the method φ_p such that $\sigma(\varphi_p,h) \le e^{-\alpha}$, then

(3.14) $$h_U(p,\alpha) \le h(p,\alpha) \le h_L(p,\alpha). \quad \blacksquare$$

Next, we consider the problem of optimality. Define the optimal complexity by

(3.15) $$C^*(\alpha) := \inf \{C(p,\alpha): \varphi_p \in \Phi\}.$$

We are interested in bounds for $C^*(\alpha)$. These are derived in

Lemma 3.1: Let f_L and f_U satisfy (2.10) and (2.11), and suppose that f_L and f_U respectively yield (via (2.13)) G_L and G_U satisfying (2.15). Then G_L and G_U have respective inverse functions p_L^*, p_U^*: $\mathbb{R}^{++} \to \mathbb{R}^{++}$ such that for all $p \in \mathbb{R}^{++}$,

(3.16) $$C_L^*(\alpha) := C_L(p_L^*(\alpha),\alpha) \le C_L(p,\alpha)$$

and

(3.17) $$C_U^*(\alpha) := C_U(p_U^*(\alpha),\alpha) \le C_U(p,\alpha),$$

with equality in (3.16) (respectively, (3.17)) if and only if $p = p_L^*(\alpha)$ (respectively, $p = p_U^*(\alpha)$).

Proof: This is an immediate corollary of Theorem 2.1. $\blacksquare$

We call $p_L^*(\alpha)$ (respectively, $p_U^*(\alpha)$) the lower (upper) optimal order, $C_L^*(\alpha)$ (respectively, $C_U^*(\alpha)$) the lower (upper) optimal complexity, and

(3.18) $h_L^*(\alpha) := h_L(p_L^*(\alpha),\alpha)$ (respectively, $h_U^*(\alpha) := h_U(p_U^*(\alpha),\alpha)$)

the lower (upper) optimal step-size. Combining (3.13), (3.15), and Lemma 3.1, we have

Theorem 3.2: Let f_L and f_U be as in Theorem 3.1. Then

$$C_L^*(\alpha) \le C^*(\alpha) \le C_U^*(\alpha). \quad \blacksquare$$

Note that if we define $p^*(\alpha)$ by

$$C(p^*(\alpha),\alpha) = C^*(\alpha),$$

we can make no statement relating $p^*(\alpha)$, $p_L^*(\alpha)$, and $p_U^*(\alpha)$. This is because we only have bounds for $C(p,\alpha)$; we do not know $C(p,\alpha)$ itself. In fact, it is important to realize what $p_L^*(\alpha)$ and $p_U^*(\alpha)$ tell us. First, consider $p_U^*(\alpha)$. We can achieve a complexity of $C_U^*(\alpha)$ by using a step-size of $h_U^*(\alpha)$, along with the method of order $p_U^*(\alpha)$. This will give optimal complexity within the sequence of algorithms for computing Φ, with cost per step of φ_p given by $c_U(p)$. Next, consider $p_L^*(\alpha)$. It is of perhaps theoretical rather than computational interest, in that we cannot compute with it. What <u>does</u> interest us is $C_L^*(\alpha)$, since it limits the theoretical improvement in $C_U^*(\alpha)$. Thus, we are interested in $p_L^*(\alpha)$ solely as a means of computing $C_L^*(\alpha)$.

We now consider behavior of these quantities as α increases and tends to infinity.

<u>Theorem 3.3</u>: Let f_L and f_U be as in Theorem 3.1. Then

(1.) $p_L^*(\alpha)$, $p_U^*(\alpha)$, $C_L^*(\alpha)$, and $C_U^*(\alpha)$ increase monotonically and tend to infinity with α.

(2.) If there exists an $M_U > 0$ such that $\kappa_U(p)^{1/p} \leq M_U$ for all p, then $\liminf_{\alpha\uparrow\infty} h_U^*(\alpha) > 0$ if $\alpha/p_U^*(\alpha)$ is bounded as $\alpha\uparrow\infty$.

(3.) If there exists an $M_L > 0$ such that $\kappa_L(p)^{1/p} \geq M_L$ for all p, then $\liminf_{\alpha\uparrow\infty} h_L^*(\alpha) > 0$ only if $\alpha/p_L^*(\alpha)$ is bounded as $\alpha\uparrow\infty$.

<u>Proof</u>: To prove (1.), it suffices to apply (1.) and (2.) of Theorem 2.2 to p_L^* and C_L^*, and to p_U^* and C_U^*. The proof of (2.) and (3.) is similar to the proof of (3.) in Theorem 2.2. ∎

Note that (1.) in Theorem 3.3 does <u>not</u> state how $p^*(\alpha)$ varies with α; as we have pointed out above, no statement about $p^*(\alpha)$ may be obtained from the information available. However, it is easy to see that $C^*(\alpha)$ increases monotonically with α and that $\lim_{\alpha\uparrow\infty} C^*(\alpha) = +\infty$.

Thus, we have extended the optimality theory of Section 2 to a more realistic situation. In Sections 5 and 6, the techniques of this section will be applied to some important basic sequences of one-step methods; we will see that the conclusions of Lemma 3.1 and Theorems 3.2 and 3.3 hold for these basic sequences.

Section 4

Normality and Order-Convergence

Let Φ be a basic sequence with respect to the error measure σ; we say that Φ is order-convergent if there exists an $h_0 > 0$ such that

$$\lim_{p\uparrow\infty} \kappa_U(p)\, h^p = 0 \quad \text{for } h \le h_0 \,. \tag{4.1}$$

Clearly, the order convergence of Φ implies that $\lim_{p\uparrow\infty} \sigma(\varphi_p,h) = 0$ for $h \le h_0$. We use the term "order-convergence" rather than "convergence," since the latter term appears extensively in the literature (e.g., Henrici [62]) and is always used to mean a "step-size convergence," i.e., $\lim_{h\downarrow 0} \sigma(\varphi,h) = 0$ for a fixed method φ.

It is intuitively plausible that as the order of an approximation increases, the approximation should improve, especially when one is trying to approximate a very smooth function. Unfortunately, Gear [71] points out that an increase in order need not always decrease the error. This situation appears in other situations in numerical mathematics; for instance, the family of Newton-Cotes quadrature formulae is not order-convergent. But suppose there exists a step-size $h_0 > 0$ for which the upper-bound error is exponentially bounded for p sufficiently large; that is, there exists $A > 0$ and $p_0 \in \mathbb{Z}^{++}$ such that

$$\kappa_U(p)\, h_0^p \le A^p \quad \text{for } p > p_0 \,. \tag{4.2}$$

If we define

$$M_U := \max \{ \max_{1\le p\le p_0} \{\kappa_U(p)^{1/p}\}, Ah_0^{-1} \},$$

we then have

$$\sigma(\varphi_p,h) \le (M_U h)^p \quad \text{for } h \le h_0 \,, p \in \mathbb{Z}^{++} \,. \tag{4.3}$$

Note that the bound in (4.3) is similar to that given by Cauchy's Integral Theorem (Ahlfors [66], pg. 122) on the normalized derivatives of an analytic function. In fact, for several classes of methods, the bound (4.3) holds whenever the solution of (2.1) is analytic.

We also formalize a weakened version of (4.3), which will be important in our study of one-step methods. Let Φ be a basic sequence, and suppose that for each $(x_0,v) \in \mathcal{D}\times\mathcal{V}$, there is a sequence $\{h_p:\ p \in \mathbb{Z}^{++}\} \subset I$ and a positive constant M_U such that

$$\sigma(\varphi_p;h) \leq (M_U h)^p \quad \text{if } h \leq h_p\ ; \tag{4.4}$$

then Φ is said to be normal. Note that (4.3) implies (4.4), while (4.4) implies (4.3) only when the sequence $\{h_p\}$ has non-vanishing support:

$$h_\Phi := \liminf_{p\uparrow\infty} h_p > 0\ . \tag{4.5}$$

If $h_\Phi = 0$, normality gives an exponential upper bound on the sequence of principal error functions (Section 3.3-5 of Henrici [62]), which are an asymptotic measure of the error as $h \downarrow 0$.

There is a simple relation between normality and order-convergence.

Proposition 4.1: Φ is order-convergent if and only if Φ is normal with nonvanishing support.

Proof: If (4.1) holds, then (in particular) we have $\lim_{p\uparrow\infty} \kappa_U(p)\, h_0^p = 0$, so that $\kappa_U(p)\, h_0^p \leq 1$ for p sufficiently large; i.e., (4.2) holds with $A = 1$. Then (as in the discussion above) (4.3) holds, implying normality with finite support.

Conversely, if (4.4) holds with finite support, we pick a positive h_0 which is less than

$$\eta := \min \{M_U^{-1}, \inf \{h_p:\ p \in \mathbb{Z}^{++}\}\}\ .$$

(Note that $\eta > 0$ by (4.5).) Let $h \le h_0$ be given, so that for some δ with $0 < \delta < 1$, we have $h = (1 - \delta)\eta$; if we define κ_U by

$$\kappa_U(p) := M_U^p ,$$

we find (since $h < h_p$) that

$$\sigma(\varphi_p, h) \le (M_U h)^p .$$

Thus

$$\kappa_U(p)\, h^p = (M_U h)^p = (M_U (1 - \delta)\eta)^p \le (1 - \delta)^p$$

(the last step since $\eta \le M_U^{-1}$), so that (4.1) holds. ∎

We are now interested in normality and order-convergence for a specific error measure σ; we will be interested in σ_{LU}, σ_L, and σ_G, which are (respectively) defined to be the maximum local error per unit step, local error per step, and global error per step over the grid. It is easy to see that a normal (order-convergent) sequence $\Phi = \{\varphi_p: p \in \mathbb{Z}^{++}\}$ with respect to σ_L naturally yields a normal (order-convergent) sequence $\Psi = \{\psi_p: p \in \mathbb{Z}^{++}\}$ with respect to σ_{LU} by setting $\psi_p := \varphi_{p+1}$ for $p \in \mathbb{Z}^{++}$. We now look at the relationships between σ_{LU} and σ_G.

Proposition 4.2: Let v have Lipschitz constant K on $\mathbb{R}^N$, and let Φ be normal (respectively, order-convergent) with respect to σ_{LU}, with M_U in (4.4) independent of $x_0 \in$ domain(v). Then Φ is normal (respectively, order-convergent) with respect to σ_G.

Proof: Let ρ be the exact relative increment function of (2.1) (as defined in Henrici [62]), so that

$$x(t_{i+1}) = x(t_i) + h\, \rho(x(t_i), h) .$$

Subtract (2.2) (with φ replaced by φ_p) from the above to get

$$e_{i+1} = e_i + h\, [\rho(x(t_i), h) - \varphi_p(x_i, h)] ,$$

where $e_i := x(t_i) - x_i$ for $0 \le i \le n$. Thus

$$\|e_{i+1}\| \le \|e_i\| + h\,\|\rho(x(t_i),h) - \rho(x_i,h)\| + h\,\|\rho(x_i,h) - \varphi_p(x_i,h)\|$$

$$\le (1 + hK)\,\|e_i\| + M_U^p\, h^{p+1} \quad \text{if } h \le h_p\ ;$$

this last step follows from the Lipschitz condition and the "uniform" normality with respect to σ_{LU}. By Lemma 1.2 of Henrici [62] and the condition $e_0 = 0$, we have

$$\|e_i\| \le K^{-1}\,[(1 + hK)^i - 1]\,(M_U h)^p$$

$$\le K^{-1}\,[(1 + hK)^n - 1]\,(M_U h)^p$$

$$\le K^{-1}\,(e^K - 1)\,(M_U h)^p$$

for all i; this gives

$$\sigma_G(\varphi_p,h) \le K^{-1}\,(e^K - 1)\,(M_U h)^p \le (Mh)^p \quad \text{if } h \le h_p,$$

for a suitably-defined $M > 0$. This proves the normality part; the remainder of the result follows from Proposition 4.1. ∎

If it is undesirable to use the "uniform normality" (i.e., the condition that M_U be independent of $x_0 \in$ domain(v) in (4.4)), we may use the following result.

Proposition 4.3: Let v be Lipschitz continuous, let Φ be normal (respectively, order-convergent) with respect to σ_{LU}, and suppose that there exists a $\lambda > 0$ such that for all $\varphi_p \in \Phi$ and all $x, y \in \mathbb{R}^N$,

$$\|\varphi_p(x) - \varphi_p(y)\| \le \lambda\, p\, \|x - y\|\ .$$

Then Φ is normal (respectively, order-convergent) with respect to σ_G.

Proof: Immediate from Theorem 3.3 of Henrici [62]. ∎

Thus normality for σ_G follows from normality for σ_{LU}, a Lipschitz condition on v and the elements of Φ, and a linear upper bound on the Lipschitz constants for the elements of Φ.

We now discuss the problem of finding uniform lower bounds on the error which are similar to the uniform upper bounds which normality provides. This will amount to

restriction of the admissible problem class $\mathcal{D}\times\mathcal{V}$ so as to guarantee that the problems are "sufficiently difficult." However, this restriction may be abandoned if we are interested only in upper bounds. We shall assume throughout the rest of this paper that there is an $M_L > 0$ (which will generally depend on Φ, σ, and the problem (x_0,v)) such that

$$\sigma(\varphi_p,h) \geq (M_L h)^p \quad \text{for } h \in I . \tag{4.6}$$

Note that (4.6) will hold for any situation in which there is no order-convergence, or in which the order-convergence (if any) is no faster than an exponential decay; moreover, in the methods we consider in Sections 5 and 6, (4.6) is a consequence of the assumption that all derivatives assume the (sharp) worst-case upper bound provided by Cauchy's estimate. It is clear that if (4.6) holds for σ_L, it holds for σ_{LU}; if (4.6) holds for σ_{LU} and if the matrix $\nabla\rho$ has only non-negative entries (with at least one positive entry), then (4.6) holds for σ_G.

It is possible to present a simplified version of the expressions derived in Section 3, under the assumption that Φ is order-convergent. We first look at the complexity of a single method within an order-convergent basic sequence.

Theorem 4.1: Let Φ be order-convergent with respect to σ. Then

$$C_L(p,\alpha) \leq C(p,\alpha) \leq C_U(p,\alpha) ,$$

where

$$C_L(p,\alpha) := M_L\, c_L(p)\, e^{\alpha/p} \quad \text{and} \quad C_U(p,\alpha) := M_U\, c_U(p)\, e^{\alpha/p} .$$

Proof: This is an immediate corollary of Theorem 3.1 and the definition of order-convergence. ∎

We may now do the optimality theory of Section 3, finding that

$$G_L(p) = p^2\, c_L'(p)/c_L(p) \quad \text{and} \quad G_U(p) = p^2\, c_U'(p)/c_U(p) . \tag{4.7}$$

Note that the assumptions (2.10) and (2.11) now state that $c_L(p)$ and $c_U(p)$ must be positive for $p > 0$ and tend to infinity with p, which is a natural way to expect the cost per step to behave. The results stated in Theorems 3.2 and 3.3 hold as before. Moreover, it should be noted that the M_U and M_L needed in (2.) and (3.) in the statement of Theorem 3.3 are precisely the M_U and M_L in (4.4) and (4.6). Thus $\liminf_{\alpha\uparrow\infty} h_U^*(\alpha) > 0$ if $\alpha/p_U^*(\alpha)$ is bounded as $\alpha \uparrow \infty$, and $\alpha/p_L^*(\alpha)$ is bounded as $\alpha \uparrow \infty$ if $\liminf_{\alpha\uparrow\infty} h_L^*(\alpha) > 0$.

Thus, the order-convergence of a basic sequence is useful in simplifying the analysis of its complexity. Of the three basic sequences we will study in this paper, two are known to be order-convergent. The proof of the order-convergence of the class of Taylor series methods is a simple consequence of the Cauchy estimate; that of the order-convergence of the (non-optimally ordered) nonlinear Brent-Runge-Kutta methods (given in Appendix B) invcives using some classical results on orthogonal polynomials to sharpen the proof in Brent [74] . (We note that it is not known whether the optimally-ordered nonlinear Brent-Runge-Kutta methods are order-convergent; it does appear likely that they are normal with vanishing support. However, we do not pursue this class of methods, because of their high combinatory cost, as indicated in Section 6.)

It is not known whether the linear Runge-Kutta methods found in Cooper [69] and in Cooper and Verner [72] are order-convergent; the best result known is the $(M_U \log(p+e))^p$ result given in Appendix A, which involves strengthening the original proof with other estimates from the theory of orthogonal polynomials. But it should be pointed out that there does exist a class of order-convergent linear Runge-Kutta methods; this is the sequence given by using the weights and abscissae for Gauss

quadrature in the methods defined on page 144 of Stetter [73]. The problem with this class of methods is that each step of φ_p requires $2^{-p}(p+1)!$ function evaluations; the prohibitive cost per step outweighs by far any advantage to be gained from the order-convergence. Thus, the question of whether there exist any order convergent linear Runge-Kutta methods which are more efficient (i.e., have smaller cost per step) remains open.

Section 5

Applications to Systems of Differential Equations

In the next two sections, we apply the theory developed in the preceeding sections to two of the most commonly-used classes of methods, i.e., Taylor series methods and Runge-Kutta methods. In Section 5, we shall treat the complexity of systems of differential equations, i.e., problems of the form (2.1) for which v is an operator on $\mathbb{R}^N$, where N is an arbitrary positive integer. In Section 6, we shall restrict our attention to the scalar case, i.e., the case where $\mathcal{V}$ consists of functions $v: \mathbb{R} \to \mathbb{R}$; for this case, Brent [74] has discovered a class of "nonlinear Runge-Kutta methods."

Before discussing the complexity of these basic sequences, we fix our error and cost measures. For the sake of definiteness, we shall choose e_G as our error measure; that is, we willl be interested in the global error, rather than the local error per step or per unit step. However, the other error measures may be used with a slight modification of the discussion contained in the sequel.

We now make precise our notion of cost. We will be concerned with the total number of arithmetic operations required. Let Φ be a given basic sequence. As in Traub and Woźniakowski [76], we shall express the cost per step associated with φ_p in the form

$$c(p) := e(\mathfrak{N}_p(v)) + d(p) \quad . \tag{5.0.1}$$

Here $\mathfrak{N}_p(v)$ is the information about v required to perform one step of φ_p, and we write $e(\mathfrak{N}_p(v))$ for the informational cost of φ_p; we call $d(p)$ the combinatory cost of φ_p.

Note that we explicitly indicate the dependence of $\mathfrak{N}_p$ on v, so that we may compare the cost of (say) an evaluation of v with a scalar arithmetic operation. Basically, $e(\mathfrak{N}_p(v))$ measures the cost of getting new data about v required by φ_p while d(p) measures the cost of combining this new data to get an approximate value of the solution at a new point. For example, Euler's method in $\mathbb{R}^N$

$$x_{i+1} = x_i + hv(x_i)$$

has informational cost $\Sigma_{i=1}^{N} e(v_i)$, where $v_1, \ldots, v_N$ are the components of v and for any function ω: $\mathbb{R}^N \to \mathbb{R}$, we define

(5.0.2) $\quad e(\omega) :=$ cost of evaluating ω at one point .

The combinatory cost is 2N arithmetic operations, i.e., one scalar multiplication and one scalar addition for each of the N components.

Finally, we now assume that $\mathcal{D}$ and $\mathcal{V}$ have been chosen so as to guarantee that the solution of (2.1) is <u>analytic</u> on I. Thus Cauchy's Integral Theorem guarantees the existence of a positive M such that for all positive integers p, we have

$$(1/p!)\, \|x^{(p)}(t)\| \le M^p \quad \text{for } t \in I .$$

5.1. Taylor Series Methods

The class Φ_T of Taylor series methods is defined by expanding x in a truncated Taylor series. Thus the increment function φ_p is given by

$$\varphi_p(x_i,h) := \sum_{k=0}^{p-1} v^{(k)}(x_i)\, h^k / (k+1)! \,, \tag{5.1.1}$$

where

$$v^{(k)}(x_i) := (d/dt)^k [v(x(t))] \Big|_{x(t) = x_i} \,. \tag{5.1.2}$$

The usual method of computing (5.1.2), as described in "classical" numerical analysis texts such as Henrici [62], invokes the chain rule. This quickly leads to expressions of horrifying complexity; for this reason, most texts quickly abandon the discussion of high-order Taylor series methods.

We are interested in faster algorithms for computing φ_p. First, we address the problem of a lower bound for the combinatory cost d(p).

Proposition 5.1.1: There exists a constant $a_L > 0$ such that any sequence of algorithms for computing Φ_T must satisfy

$$d(p) \geq a_L\, p^N \,. \tag{5.1.3}$$

Proof: Any algorithm for computing φ_p requires the information

$$\mathfrak{N}_p(v) := \{D^\beta v:\ 0 \leq |\beta| \leq p - 1\} \,.$$

(We use the standard multi-index notation found in Friedman [69].) It is then easy to see that the above set has $O(p^N)$ (as $p \uparrow \infty$) distinct elements, which are (generally) independent; this is an immediate consequence of Problem 11 in Chapter I of Pólya and Szegő [25]. Thus (5.1.3) gives a linear lower bound. ∎

Note that the constant a_L in (5.1.3) depends on N. Since we are treating the case where N is fixed and p is allowed to very, we will not indicate this dependence explicitly. We now see how close we can get to an optimum value for d(p).

Theorem 5.1.1: There exists a constant $a_U > 0$ such that the combinatory cost $d(p)$ of computing $\varphi_p \in \Phi_T$ satisfies the bound

$$d(p) \leq a_U p^N \ln (p+e) . \tag{5.1.4}$$

Proof: We first consider the case $N = 1$. Note that $x(h)$ is the zero of

$$F(z) := \int_{x_0}^{z} d\xi / v(\xi) - h . \tag{5.1.5}$$

As in Brent and Kung [76], we consider the formal power series

$$P(s) := F(x_0+s) - F(x_0) ,$$

where s is an indeterminate. Let V be the power series reversion of P. Adopting the notation of Brent and Kung [76], we see that

$$x(s) = x_0 + V(s) = x_0 + V_p(s) + O(s^{p+1}) .$$

By the uniqueness of the Taylor coefficients of an analytic function, we see that

$$\varphi_p(x_0,h) = h^{-1} V_p(h) .$$

Since the number $V_p(h)$ can be computed in $O(p \ln p)$ operations from the Taylor coefficients of v (by Theorem 6.2 of Brent and Kung [76]), the result for $N = 1$ follows.

For $N \geq 2$, we use Newton's method (Rall [69]) applied to the formal power series operator P given by

$$(Py)(s) := y(s) - x_0 - \int_0^s v(y(\tau))\, d\tau ;$$

clearly, the formal power series $x(s)$ is the zero of P. The algorithm itself is defined recursively. Let a formal power series $x_{(p)}(s)$ satisfying

$$x_{(p)}(s) = x(s) + O(s^{p+1})$$

be given. Precompute

$$w(s) := \int_0^s v(x_{(p)}(\tau))\, d\tau - x_0 - x_{(p)}(s) + O(s^{2p+2}) , \tag{5.1.6}$$

$$Q(s) := \nabla v(x_{(p)}(s)) + O(s^{2p+2}) , \tag{5.1.7}$$

and let $u_{(0)}(s) := 0$. Then set

$$x_{(2p+1)}(s) := x_{(p)}(s) + u_{(p+1)}(s) ,$$

where

$$(5.1.8)\quad u_{(k+1)}(s) := \int_0^s Q(\tau)\, u_{(k)}(\tau)\, d\tau + w(s) + O(s^{2p+2}),\quad 0 \le k \le p\ .$$

Following the proof given in Rall [69], we find that

$$x_{(2p+1)}(s) = x(s) + O(s^{2p+2})\ .$$

We need only consider the cost T(p,N) of computing the series $x_{(p)}(s)$ in determining d(p), since x(h) may be recovered from the formal power series in O(p) operations. Clearly, we have the recursion

$$(5.1.9)\qquad T(2p+1,N) \le T(p,N) + T_6 + T_7 + T_8\ ,$$

where T_m is the cost of step (5.1.m) for m = 6, 7, 8. Let COMP(p,N) be the time required to find the first p terms of the formal power series $f(y_1(s), \ldots, y_N(s))$, where f, $y_1, \ldots, y_N$ are formal power series, and $y_1, \ldots, y_N$ have zero constant term. Theorem 7.1 of Brent and Kung [76] states that

$$COMP(p,2) = O(p^2 \ln p)\ ,$$

and it is easy to show that for any $N \in \mathbb{Z}^{++}$,

$$COMP(p,N+1) = O(p\ COMP(p,N))\ .$$

Thus for $N \ge 2$, we have

$$(5.1.10)\qquad COMP(p,N) = O(p^N \ln p)\ ,$$

and so we see that

$$T_6 + T_7 = O((2p+1)^N \ln p)\ .$$

Finally, let MULT(p) be as in Brent and Kung [76]; we see that

$$T_8 = (p+1)\,[N^2\, MULT(2p+1) + O(p)] = O((2p+1)^2 \ln p)$$

if Fast Fourier Transform multiplication (Borodin and Munro [75]) is used. Since $N \ge 2$, we have

$$(5.1.11)\qquad T_6 + T_7 + T_8 = O((2p+1)^N \ln p)\ ,$$

and so (5.1.9) and (5.1.11) imply that

$$T(p,N) = O(p^N \ln p),$$

which completes the proof. ∎

(Note that the second algorithm is inferior to the first algorithm when applied to the scalar case N = 1, where we find that the second algorithm requires $O(p^2 \ln p)$ arithmetic operations.)

We now determine bounds on $C(p,\alpha)$. First, consider lower bounds. Clearly, there exists $e_L(v) \geq 0$ such that

$$e(D^\beta v_i) \geq e_L(v) \quad (1 \leq i \leq n, |\beta| \in \mathbb{Z}^+) . \tag{5.1.12}$$

Since $\mathfrak{N}_p(v)$ has $O(p^N)$ elements, there exists a constant $b_L > 0$ such that

$$e(\mathfrak{N}_p(v)) \geq b_L\, e_L(v)\, p^N . \tag{5.1.13}$$

From (5.1.3) and (5.1.13), we have a lower-bound cost per step of

$$c_L(p) = [a_L + b_L\, e_L(v)]\, p^N . \tag{5.1.14}$$

This leads to

Theorem 5.1.2: $C_L(p,\alpha) = M_L\, [a_L + b_L\, e_L(v)]\, p^N\, e^{\alpha/p}$.

Proof: This is an immediate consequence of (4.6) and (5.1.14). ∎

Note that $f_L(p) := M_L c_L(p)$ satisfies the conditions of Theorem 3.2. So the optimality theory of Section 3 holds. In particular, we have

Theorem 5.1.3: $C_L^*(\alpha) = M_L\, [a_L + b_L\, e_L(v)]\, (e/N)^N\, \alpha^N$.

Proof: From $(4.7)_1$ and (5.1.14), we find that $G_L(p) = Np$, so that

$$p_L^*(\alpha) = \alpha/N \quad \text{and} \quad h_L^*(\alpha) = (M_L e^N)^{-1} .$$

The result follows by letting $p = p_L^*(\alpha)$ in the definition of $C_L(p,\alpha)$. ∎

However, recall that we assumed that the non-identical mixed partial derivatives of v are independent. There are a number of systems for which this is not true (for instance, constant coefficient linear systems); for such systems, it is clear that we may

be able to use the extra information of non-independence to find algorithms that are faster than the lower bounds given above. However, we will ignore this case and only consider the problem for a "general" function v.

Next, we turn to upper bounds on the complexity. Theorem 5.1.1 tells us how to combine the necessary information to get the solution at a new grid-point; we need only measure the cost of getting the information. So, let

$$e^{(k)}(v) = \max \{e(D^{\beta}v_i): 1 \le i \le N, |\beta| = k\} .$$

Using the result in Pólya and Szegö [25], we see that

$$e(\mathfrak{N}_p(v)) \le N \Sigma_{k=0}^{p-1} e^{(k)}(v) (N+k-1)! / [k!(N-1)!] . \tag{5.1.15}$$

Unfortunately, the right-hand side of (5.1.15) does not fit our general model, so we must assume that we know how $e^{(k)}(v)$ changes as k increases. We will consider the case where the cost of derivative evaluation is bounded; that is, we will assume that

$$e^{(k)}(v) \le e_U(v) \tag{5.1.16}$$

for some $e_U(v)$ independent of k. Other cases (e.g., $e^{(k)}(v) = O(k^m)$ for some $m > 0$) may be analyzed in a similar manner; of course, they will give different results. By (5.1.15) and (5.1.16), there is a $b_U > 0$ such that

$$e(\mathfrak{N}_p(v)) \le b_U e_U(v) p^N . \tag{5.1.17}$$

From (5.1.4) and (5.1.17), we have an upper-bound cost per step of

$$C_U(p) = a_U p^N \ln (p+e) + b_U e_U(v) p^N . \tag{5.1.18}$$

This leads to

Theorem 5.1.4: There exists an $M_U > 0$ such that

$$C_U(p,\alpha) = M_U [a_U p^N \ln (p+e) + b_U e_U(v) p^N] e^{\alpha/p} .$$

Proof: By Cauchy's Integral Theorem, there exists a $B > 0$ such that

$$|||x^{(k+1)}||| / (k+1)! \le B^k ,$$

where we define

(5.1.19) $$|||y||| := \max_{t \in I} \|y(t)\|$$

for any $y: I \to \mathbb{R}^N$. Thus by Section 3.3-3 of Henrici [62], we see that a Lipschitz constant for φ_p in Φ_T is given by

$$\Sigma_{k=0}^{p-1} |||x^{(k+1)}||| \, h^k / (k+1)! \;\le\; \Sigma_{k=0}^{p-1} (Bh)^k \;\le\; L := (1 - Bh_0)^{-1},$$

provided that $h \le h_0 < B^{-1}$. By Section 3.3-4 of Henrici [62] and Proposition 4.3, there exists an $M_U > 0$ such that

$$\sigma_G(\varphi_p, h) \;\le\; (M_U h)^p .$$

The result now follows from Theorem 4.1 and (5.1.18). ∎

We are now ready to consider the optimal p for $C_U(p,\alpha)$.

Theorem 5.1.5:

(1.) For all $\alpha > 0$, there exists $p_U^*(\alpha)$ such that (3.17) holds.

(2.) $p_U^*(\alpha)$ increases monotonically with α, and

$$p_U^*(\alpha) \sim \alpha/N \quad \text{as } \alpha \uparrow \infty .$$

(3.) $C_U^*(\alpha)$ increases monotonically with α, and

$$C_U^*(\alpha) \sim M_U a_U (e/N)^N \alpha^N \ln \alpha \quad \text{as } \alpha \uparrow \infty .$$

(4.) $h_U^*(\alpha) \sim (M_U e^N)^{-1}$ as $\alpha \uparrow \infty$.

Proof: Clearly c_U satisfies (2.10) and (2.11). Now write

$$G_U(p) = G_1(p) + G_2(p),$$

where

$$G_1(p) = N p \quad \text{and} \quad G_2(p) = \nu p^2 / D_2(p);$$

here we set

$$D_2(p) := (p+e)[(p+e)\ln(p+e) + 1] \quad \text{and} \quad \nu := a_U / [b_U e_U(v)] .$$

We see immediately that G_1 satisfies (2.15); a straightforward calculation shows that

$$G_2'(p) = \nu [D(p)]^{-2} \{\nu p [\ln(p+e)] - 1] + 2e[\nu \ln(p+e) + 1]\},$$

so that $G_2'(p) > 0$ for $p > 0$. Thus G_2 satisfies (2.15), which shows that G_U satisfies (2.15). Hence p_U^* and C_U^* behave as described in Theorem 3.3.

Since $p_U^*(\alpha)$ goes to infinity with α, we see that

$$\alpha = G_U(p_U^*(\alpha)) \sim N\, p_U^*(\alpha) + p_U^*(\alpha) / \ln p_U^*(\alpha) \sim N\, p_U^*(\alpha),$$

which gives the asymptotic estimate in (2.). The rest of the Theorem follows from this estimate. ∎

Unfortunately, the estimates given above are only asymptotic as $\alpha \uparrow \infty$; this will be typical, since many of the equations to be solved involve products of logarithmic and polynomial terms, and thus cannot be solved exactly. On the other hand, these asymptotic expressions are sufficient for our purposes, since they describe how quickly $p_U^*(\alpha)$ and $C_U^*(\alpha)$ increase with α.

Note that as α tends to infinity, $C_U^*(\alpha)$ becomes independent of $e_U(v)$, which measures how hard it is to evaluate the derivatives of v; this is because the combinatory cost eventually overwhelms the informational cost. This kind of behavior will be typical of the complexity analyses in this paper. Finally, note that the bound

$$(5.1.20)\quad C_L^*(\alpha) = O(\alpha^N) \le C^*(\alpha) \le O(\alpha^N \ln \alpha) = C_U^*(\alpha) \text{ as } \alpha \uparrow \infty$$

implies that

$$C_U^*(\alpha) / C_L^*(\alpha) = O(\ln \alpha) \text{ as } \alpha \uparrow \infty;$$

this indicates the gap in our knowledge of the complexity of solving (2.1) via Taylor series methods.

5.2 Linear Runge-Kutta Methods

For many functions v, caculation of the derivatives required by Taylor series methods is prohibitively expensive. For this reason, we are interested in methods which use information that is somewhat more readily available. In particular, we will consider methods that use only evaluations of v, combined in a highly structured manner. We say that Φ_{LRK} is a class of linear Runge-Kutta methods (abbreviated, "LRK methods") if each increment function φ_p may be written in the form

$$\varphi_p(x_i,h) := \Sigma_{l=0}^{s-1} \lambda_{sl} k_l \tag{5.2.1}$$

where

$$k_l := v(x_i + h \Sigma_{j=0}^{l-1} \lambda_{lj} k_j) \quad \text{for } 0 \le l \le s-1, \tag{5.2.2}$$

the integer $s = s(p)$ is said to be the number of stages of φ_p; the number of stages is equal to the number of times the vector function v must be evaluated. (In order to simplify notation, we will not explicitly indicate the dependence of λ_{lj} and k_j on p.) The method φ_p defined by (5.2.1) and (5.2.2) is explicit in that k_l depends only on $k_0, \dots, k_{l-1}$; see Butcher [64a] for a discussion of semi-explicit and implicit methods.

Since the function φ_p is (in practice) always evaluated by using the obvious algorithm suggested by its definition, we shall identify an algorithm for evaluating φ_p with φ_p itself. Thus the problem of finding the best algorithm for evaluating φ_p in Φ_{LRK} is equivalent to the problem of finding the best basic sequence of LRK methods possible. This is related to the problem of finding the smallest value of s(p) such that φ_p has order p. This minimal value is given by

$$(5.2.3) \qquad s(p) = \begin{cases} p & p = 1, 2, 3, 4 \\ p+1 & p = 5, 6 \\ p+2 & p = 7 \\ \text{unknown} & p \geq 8 \end{cases} .$$

For methods of order greater than seven, a gap develops. For instance, eighth-order methods with eleven stages exist, and it is known that any eighth-order method requires at least ten stages. For arbitrary $p \geq 8$, the best bounds known for the optimum value of $s(p)$ are

$$(5.2.4) \qquad p + \vartheta(p) \leq s(p) \leq (p^2 - 7p + 14) / 2 ,$$

where $\vartheta(p) \geq c \ln p$ for all p sufficiently large (for some $c > 0$). The lower bound is given in Butcher [75]; the proof is quite involved, and the result is not much better than the "trivial" lower bound $s(p) \geq p$ (Hindmarsh [74], page 84). A class Φ_{CVRK} of methods such that φ_p requires only $(p^2 - 7p + 14) / 2$ stages is given in Cooper and Verner [72].

We first consider lower bounds on the complexity $C(p,\alpha)$ using LRK methods. The "trivial" lower bound $s(p) \geq p$ will be used, since the term $\vartheta(p)$ will be small when p is small and will not affect the asymptotic behavior of optimal order and complexity for p large. It is known (Butcher [64]) that at least $O(p^2)$ of the subdiagonal elements of the matrix A (whose elements are the λ_{ij} in (5.2.2)) must be non-zero in order for A to define a p^{th}-order method. Thus there exists $a_L > 0$ such that

$$(5.2.5) \qquad d(p) \geq a_L p^2 ;$$

since $s(p) \geq p$, we see that

$$(5.2.6) \qquad e(\mathfrak{N}_p(v)) \geq N e_L(v) p ,$$

where we now write

$$e_L(v) := \min_{1 \leq i \leq N} e(v_i) .$$

Thus (5.2.5) and (5.2.6) show that a lower bound on the cost per step for φ_p is given by

$$c_L(p) = a_L p^2 + N e_L(v)\, p \quad . \tag{5.2.7}$$

Theorem 5.2.1:

$$C_L(p,\alpha) = M_L [a_L p^2 + N e_L(v)\, p]\, e^{\alpha/p} \quad .$$

Proof: This follows immediately from (4.6) and (5.2.7). ∎

It is clear that $f_L(p) := M_L [a_L p^2 + N e_L(v)\, p]\, e^{\alpha/p}$ satisfies (2.10) and (2.11). We claim that f_L yields a G_L satisfying (2.15). Indeed, write

$$f_L(p) = f_1(p)\, f_2(p) ,$$

where

$$f_1(p) := M_L a_L p$$

and

$$f_2(p) := p + \nu , \quad \text{where} \quad \nu := N e_L(v) / a_L \quad .$$

Clearly f_1 yields a G_1 satisfying (2.15). Since f_2 is a linear polynomial with a negative zero, it may be shown that f_2 yields a G_2 satisfying (2.15). Thus f_L yields a G_L satisfying (2.15); in fact, we have

$$G_L(p) = G_1(p) + G_2(p) = p\,[1 + (1 + \nu p^{-1})^{-1}] \quad . \tag{5.2.8}$$

This leads us to

Theorem 5.2.2:

$$C_L^*(\alpha) \sim [M_L a_L e^2 / 4]\, \alpha^2 \quad \text{as} \quad \alpha \uparrow \infty \quad .$$

Proof: From (5.2.8), we see that $G_L(p) \sim 2p$ as $p \uparrow \infty$. Since (2.10), (2.11), and (2.15) hold, $p_L^*(\alpha)$ tends to infinity with α. Thus

$$\alpha = G_L(p_L^*(\alpha)) \sim 2\, p_L^*(\alpha) \quad \text{as} \quad \alpha \uparrow \infty ,$$

i.e., $p_L^*(\alpha) \sim \alpha/2$ as $\alpha \uparrow \infty$. The result now follows from Theorem 5.2.1. ∎

We now turn to upper bounds on complexity. The class Φ_{CVRK} derived in Cooper and Verner [72] has two deficiencies, the first of which is that no uniform upper bound on $e_{LU}(\varphi_p,h)$ is known for Φ_{CVRK}; in addition, the combinatory cost for this class of methods is $O(p^4)$ as $p \uparrow \infty$. Instead, we turn to the basic sequence Φ_{CRK} discussed in Appendix A. There, we prove that there is an $M_U > 0$ such that

$$e_G(\varphi_p,h) \leq (M_U \ln (p + e)\, h)^p , \tag{5.2.9}$$

provided $h \leq h_p$, where $h_p = O((\ln p)^{-1})$ as $p \uparrow \infty$. Furthermore, there are a large number of extra zeros in the matrix A for $\varphi_p \in \Phi_{CRK}$. Using the notation of Appendix A, we see that the number of non-zero entries in A is

$$\begin{aligned} \Sigma_{i=0}^{s} \xi_i &= \Sigma_{i=1}^{p-1} i^2 + p \\ &= p^3/3 - p^2/2 + 7p/6 \\ &\leq p^3/3 + 2p^2/3 \end{aligned}$$

for $p \in \mathbb{Z}^{++}$. Finally, note that the number of stages $s(p)$ required for $\varphi_p \in \Phi_{CRK}$ is

$$s(p) = \lfloor (p^2 - 2p + 4)/2 \rfloor \leq p^2/2 + p \tag{5.2.10}$$

for $p \in \mathbb{Z}^{++}$, which shows that the number of stages required for a p^{th}-order method in Φ_{CRK} asymptotically equals the number requires for a p^{th}-order method in Φ_{CVRK}. Thus (considering the combinatory costs), the class Φ_{CVRK} actually costs more per step than does Φ_{CRK}; ignoring the combinatory costs would have caused us to reach the opposite conclusion.

First, we look at the cost per step. By (5.2.10),we see that

$$e(\mathfrak{M}_p(v)) \leq \frac{1}{2} (p^2 + p)\, N\, e_U(v) , \tag{5.2.11}$$

where

$$e_U(v) := \max_{1 \leq i \leq N} e(v_i) .$$

Since we are using Φ_{CRK}, it is easy to see that there is a $b_U \geq 2/3$ such that

$$d(p) \leq (p^3/3 + b_U p^2) \cdot N . \tag{5.2.12}$$

Combining (5.2.11) and (5.2.12), we see that the total combinatory cost per step is bounded by

$$c_U(p) = N [\, 2p^3/3 + \beta_1 p^2 + \beta_2 p \,], \tag{5.2.13}$$

where

$$\beta_1 := e_U(v) / 2 + 2 b_U \quad \text{and} \quad \beta_2 := e_U(v) / 2 .$$

Using (5.2.9) and (5.2.13) gives

Theorem 5.2.3:

$$C_U(p,\alpha) = M_U N [2p^3/3 + \beta_1 p^2 + \beta_2 p \,] \ln (p + e)\, e^{\alpha/p} . \quad \blacksquare$$

Now we look at the optimality theory for the upper bound.

Theorem 5.2.4.:

(1.) For all $\alpha > 0$, there exists $p_U^*(\alpha)$ such that (3.17) holds.

(2.) $p_U^*(\alpha)$ increases monotonically with α, and

$$p_U^*(\alpha) \sim \alpha/3 \quad \text{as } \alpha \uparrow \infty .$$

(3.) $C_U^*(\alpha)$ increases monotonically with α, and

$$C_U^*(\alpha) \sim [\, 2 M_U N e^3 / 81 \,] \alpha^3 \ln \alpha \quad \text{as } \alpha \uparrow \infty.$$

(4.) $h_U^*(\alpha) \sim (M_U e^3 \ln \alpha)^{-1}$ as $\alpha \uparrow \infty$.

Proof: We write

$$f_U(p) := M_U \ln (p + e)\, c_U(p)$$

in the form

$$f_U(p) = f_1(p)\, f_2(p),$$

where

$$f_1(p) = M_U N p \ln (p + e) \quad \text{and} \quad f_2(p) = 2p^2/3 + \beta_1 p + \beta_2 .$$

As was pointed out in Section 2, f_1 satisfies the hypotheses of Theorem 2.1. Now we consider f_2. Clearly f_2 has no positive zeros; it may be seen that the condition

$b_U \geq 2/3$ implies that f_2 has a positive discriminant and hence has no complex roots. Thus f_2 has only negative roots; one may then show that this guarantees that f_2 satisfies the hypotheses of Theorem 2.1. By Lemma 2.1, the same may be said for f.

Thus p_U^* and C_U^* behave as described in (1.) of Theorem 3.3. We also see that $G_U(p) \sim 3\,p$ as $p \uparrow \infty$. Thus the estimate in (2.) holds, from which we get the estimates in (3.) and (4.). ∎

So in the class of linear Runge-Kutta methods, we find that

(5.2.14) $$C_L^*(\alpha) = O(\alpha^2) \leq C^*(\alpha) \leq C_U^*(\alpha) = O(\alpha^3 \ln \alpha)$$

as α tends to infinity; hence, the ratio

$$C_U^*(\alpha) / C_L^*(\alpha) = O(\alpha \ln \alpha)$$

indicates the gap in our knowledge of the complexity of linear Runge-Kutta methods.

Finally, we wish to compare the classes of Taylor series methods and LRK methods. Write $C_{U,T}^*$, $C_{L,T}^*$, and C_T^* (respectively, $C_{U,LRK}^*$, $C_{L,LRK}^*$, and C_{LRK}^*) for C_U^*, C_L^*, and C^* in the class Φ_T (respectively, the class Φ_{LRK}). Since we have only asymptotic expressions for these quantities, we are forced to use an asymptotic comparison. If $f, g : \mathbb{R}^{++} \to \mathbb{R}^{++}$ satisfy $\lim_{\alpha \uparrow \infty} f(\alpha) = \lim_{\alpha \uparrow \infty} g(\alpha) = +\infty$, we will write

(5.2.15) $$f \prec g \quad \text{iff} \quad f(\alpha) = o(g(\alpha)) \text{ as } \alpha \uparrow \infty ;$$

we say that f is <u>asymptotically less than</u> g. If $f \prec g$, there is an $\alpha_0 > 0$ such that $f(\alpha) < g(\alpha)$ for $\alpha > \alpha_0$, so there is a non-asymptotic interpretation of the order relation $\prec$. In addition, we see that if $f \prec g$, then $g(\alpha)$ grows much more quickly than $f(\alpha)$ does as α increases. Using the results of (5.1.20) and (5.2.14), we then have the following

Theorem 5.2.5: Suppose that (5.1.16) holds.

(1.) If $N = 1$, then $C_{U,T}^* < C_{L,LRK}^*$.

(2.) If $N = 2$, then $C_{U,T}^* < C_{U,LRK}^*$.

(3.) If $N = 3$, then

$$C_{U,T}^*(\alpha) = O(C_{U,LRK}^*(\alpha))$$

and

$$C_{U,LRK}^*(\alpha) = O(C_{U,T}^*(\alpha))$$

as $\alpha \uparrow \alpha_*$.

(4.) If $N \geq 4$, then $C_{U,LRK}^* < C_{L,T}^*$. ∎

If (5.1.16) does not hold, then (1.), (2.), and (3.) may be false, but (4.) will certainly be true. As an immediate corollary to the above theorem, we have

Theorem 5.2.6:

(1.) If $N = 1$ and (5.1.16) holds, then $C_T^* < C_{LRK}^*$.

(2.) If $N \geq 4$, then $C_{LRK}^* < C_T^*$. ∎

So if derivatives are cheap to evaluate, we see that the best Taylor series method known is better than the best linear Runge-Kutta method possible for the scalar case $N = 1$; but if $N \geq 4$, the best linear Runge-Kutta method known is better than the best Taylor series method possible.

Section 6

Nonlinear Runge-Kutta Methods for the Scalar Case

We now consider a generalization of the familiar LRK methods, based on the description in Brent [74], [76]. A basic sequence Φ_{NRK} is said to be a class of nonlinear Runge-Kutta methods (abbreviated, "NRK methods") if each increment function φ_p may be written in the form

$$\varphi_p(x_i,h) := \tau_s(x_i,h; k_0, \ldots, k_{s-1}), \tag{6.1}$$

where

$$k_j := v(y_j), \quad y_j := \tau_j(x_i,h; k_0, \ldots, k_{j-1}) \quad (0 \le j \le s-1) \tag{6.2}$$

for suitable functions $\tau_j: \mathbb{R}^N \times \mathbb{R} \times (\mathbb{R}^N)^j \to \mathbb{R}^N$ $(0 \le j \le s)$; as in the linear case, $s = s(p)$ is the number of stages (i.e., evaluations of v) of φ_p. Again, for notational convenience, we do not indicate the dependence of the k_j, y_j, and τ_j on x_i and p.

In the remainder of this section, we will only consider the scalar case $N = 1$, since it is not known whether NRK methods exist for larger values of N. In this case, (5.1.5) shows how an s-stage NRK method of order p may derived from a $(p+1)^{th}$-order iterative method for solving the nonlinear equation

$$F(z) = 0, \tag{6.3}$$

using Brent-information (Meersman [76]) of the form

$$\mathfrak{I}_{B,s}(F) := \{F(x_0), F'(x_0), F'(y_1), \ldots, F'(y_{s-1})\}. \tag{6.4}$$

Brent [74], [75] used this transformation to derive a sequence Φ_{MBRK} of (modified) Brent-Runge-Kutta methods ("BRK methods"), in which the s-stage method has order

$$p = 2s - 1. \tag{6.5}$$

Furthermore, Meersman [76] proved that this order is the greatest possible in the class of all such BRK methods. We now extend Meersman's result to include all NRK methods.

Theorem 6.1: No s-stage NRK method can have order greater than 2s - 1.

Proof: Let φ be an s-stage method with order p. We will construct (from φ) an iterative method ψ of order $q := p + 1$ for finding a simple zero ξ of an arbitrary analytic function $F : \mathbb{R} \to \mathbb{R}$.

The method ψ is defined as follows. Let x_0 be an approximation to ξ such that F' is nonzero between x_0 and ξ. (Since $F'(\xi) \neq 0$, such an x_0 exists.) Write $t_0 := F(x_0)$; without loss of generality, assume $t_0 < 0$. Now apply one step of φ, using a step-size of $-t_0$, to the problem

$$\dot{x}(t) = F'(x(t))^{-1} \quad (t_0 < t < 0) \quad \text{with} \quad x(t_0) = x_0 ,$$

whose solution is the functional inverse of F

$$x(0) = F^{-1}(0) = \xi ;$$

then ψ is given by

$$\psi(x_0) := x_0 - t_0 \varphi(x_0, -t_0) .$$

By the definition of order for iterative methods, it is clear that ψ has order q; moreover, ψ uses the generalized Brent information (Definition II.3.8 of Meersman [76])

$$\mathfrak{I}_{GB,s} := \{F(x_0), F'(y_0), F'(y_1), \ldots, F'(y_{s-1})\} .$$

Suppose that $y_0 \neq x_0$; then $q \leq 2s$ by Theorem II.3.3 of Meersman [76]. On the other hand, if $y_0 = x_0$, then ψ uses the Brent-information (6.4); by Theorem II.2.4 of Meersman [76] (also due to Woźniakowski), we have $q \leq 2s$ in this case also. Thus in either case, we find that

$$p + 1 = q \leq 2s ,$$

and the desired result follows. ∎

Thus Φ_{MBRK} is informationally-optimal in the class of NRK methods, in the sense that each φ_p in Φ_{MBRK} uses the minimum number of stages possible for a p^{th}-order NRK method.

We will now derive lower bounds for the complexity $C(p,\alpha)$ via NRK methods. Clearly, Theorem 6.1 implies that

$$e(\mathfrak{N}_p(v)) \geq e(v)\,(p+1)/2\,, \tag{6.6}$$

and a linear lower bound on the combinatory cost states that

$$d(p) \geq a_L\, p \tag{6.7}$$

for some $a_L > 0$. By (6.6) and (6.7), a lower bound on the cost per step for φ_p is

$$c_L(p) = (a_L + e(v)/2)\, p + e(v)/2\,, \tag{6.8}$$

which leads to

Theorem 6.2: $C_L(p,\alpha) = M_L\,[(a_L + e(v)/2)\, p + e(v)/2]\, e^{\alpha/p}$.

Proof: This follows immediately from (4.6) and (6.8). ∎

Note that $f_L(p) := M_L\, c_L(p)$ is a linear polynomial with a negative zero; it then follows that f_L satisfies the conditions of Theorem 3.2. So, the optimality theory of Section 3 holds; in particular, we have

Theorem 6.3: $C_L^*(\alpha) \sim M_L\, e\,[a_L + e(v)/2]\,\alpha$ as $\alpha \uparrow \infty$.

Proof: From $(4.7)_1$ and (6.8), we find that

$$G_L(p) = p^2/(p+\beta^{-1})\,, \quad \text{where} \quad \beta := 1 + 2\,a_L/e(v)\,,$$

and so $G_L(p) \sim p$ as $p \uparrow \infty$; thus $p_L^*(\alpha) \sim \alpha$ as $\alpha \uparrow \infty$. The result follows by letting $p = p_L^*(\alpha)$ in the definition of $C_L(p,\alpha)$. ∎

Next, we consider upper bounds on the number of operations required. Instead of using Φ_{MBRK}, we will use the class Φ_{BRK} of "unmodified" BRK methods described in Appendix B, where it is shown that there is an $M_U > 0$ such that

$$\sigma_G(\varphi_p,h) \leq (M_U\, h)^p\,; \tag{6.9}$$

no such bound is known for Φ_{MBRK}. In addition, Φ_{MBRK} requires the solution of $p - 1$ linear systems of equations, the i^{th} having $p - i$ unknowns, in order to perform a "reorthogonalization." So the smallest known combinatory cost for this class is about $O(p^{3.81})$ arithmetic operations; this is obtained by using Strassen's technique for linear systems (described in Borodin and Munro [75]). On the other hand, most of the combinatory cost for φ_p in Φ_{BRK} is involved in finding the coefficients of the polynomial p_{n+1} (see Appendix B); once these coefficients are known, the remaining combinatory cost is $O(p \ln p)$ as $p \uparrow \infty$. An estimate of how much work is required to compute these coefficients is given in

Lemma 6.1: Let $x_0, y_1, \ldots, y_r, w_0, z_0, \ldots, z_r$ be given, and let

$$Q(x) := \Sigma_{i=0}^{r+1} q_i x^i$$

be the unique polynomial of degree at most $r + 1$ satisfying

$$Q(x_0) = w_0, \quad Q'(x_0) = z_0, \quad \text{and} \quad Q'(y_i) = z_i \quad (1 \le i \le r) .$$

If $T(r)$ is the time required to compute $q_0, \ldots, q_{r+1}$, then

$$T(r) = O(r \ln^2 r) \text{ as } r \uparrow \infty .$$

Proof: The coefficients $q_1, 2q_2, \ldots, (r+1)q_{r+1}$ of Q' may be computed in time $O(r \ln^2 r)$ by using a fast algorithm for computing the coefficients of the Lagrange polynomial interpolating the points $(x_0, z_0), (y_1, z_1), \ldots, (y_r, z_r)$; see Borodin and Munro [75] for details. Then $O(r)$ operations yield $q_1, \ldots, q_{r+1}$, and Horner's rule gives q_0 with $O(r)$ additional operations. ∎

Thus there exists $a_U > 0$ such that

$$d(p) \le a_U p \ln^2(p+e) . \tag{6.10}$$

In order to simplify matters a bit, note that Theorem B.1 implies that

$$e(\mathfrak{R}_p(v)) \le e(v)\, p . \tag{6.11}$$

Although the estimate above is not exact for $p > 2$, it is asymptotically equal to that in Theorem 8.1. (If necessary, the sharper estimate given there may be used, but the calculation of optimal order involves considerably more detail, the results of which are not particularly enlightening.) Combining (6.10) and (6.11), we see that the cost per step is bounded by

$$c_U(p) = e(v)\, p + a_U\, p \ln^2(p+e) \ . \tag{6.12}$$

Thus (6.9) and (6.12) imply

Theorem 6.4: $C_U(p,\alpha) = M_U [e(v)\, p + a_U\, p \ln^2(p+e)]\, e^{\alpha/p}$. ∎

We now determine the behavior of $p_U^*(\alpha)$. Here we find that $f_U(p) := M_U\, c_U(p)$ may be decomposed as

$$f_U(p) = f_1(p)\, f_2(p) \ ,$$

where

$$f_1(p) := M_U\, e(v)\, p \quad \text{and} \quad f_2(p) := 1 + \beta \ln^2(p+e) \ ,$$

and $\beta := a_U / e(v)$. Clearly f_1 and f_2 satisfy (2.10) and (2.11), and f_1 yields a G_1 satisfying (2.15). We need only check that f_2 yields a G_2 satisfying (2.15). But

$$G_2(p) = 2\beta\, p^2 \ln (p+e) / D_2(p) \ , \quad \text{where} \quad D_2(p) := (p+e)\, f_2(p) \ ,$$

so that setting

$$g_2(p) := \beta\, p \ln^2(p+e) [\ln (p+e) - 1] + 2\beta\, e \ln^2(p+e) + (p + 2e) \ln (p+e) + p \ ,$$

we find that $p > 0$ implies

$$G_2'(p) = 2\beta\, p\, g_2(p) / D_2(p)^2 > 0 \ .$$

Thus the optimality theory applies.

Theorem 6.5:

(1.) $p_U^*(\alpha)$ increases monotonically with α, and

$$p_U^*(\alpha) \sim \alpha \quad \text{as } \alpha \uparrow \infty .$$

(2.) $C_U^*(\alpha)$ increases monotonically with α, and

$$C_U^*(\alpha) \sim M_U a_U e \alpha \ln^2 \alpha \quad \text{as } \alpha \uparrow \infty .$$

(3.) $h_U^*(\alpha) \sim (M_U e)^{-1}$ as $\alpha \uparrow \infty$.

Proof: (1.) follows from the fact that $G_U(p) \sim p$ as $p \uparrow \infty$; (2.) and (3.) follow from (1.) and Theorem 6.4. ∎

So in the class of nonlinear Runge-Kutta methods, we find that

$$C_L^*(\alpha) = O(\alpha) \le C^*(\alpha) \le C_U^*(\alpha) = O(\alpha \ln^2 \alpha) \tag{6.13}$$

as α tends to infinity; so, the ratio

$$C_U^*(\alpha) / C_L^*(\alpha) = O(\ln^2 \alpha) \quad \text{as } \alpha \uparrow \infty$$

indicates the gap in our knowledge of the complexity of nonlinear Runge-Kutta methods.

Finally, we wish to compare the classes of Taylor series methods and NRK methods. Adopting the notation at the end of Section 5.2 in an obvious manner, we have

Theorem 6.6: If (5.1.16) holds, then $C_{U,T}^* < C_{U,NRK}^*$.

Proof: Immediate from (5.1.20) and (6.13). ∎

Thus if derivatives of v are easy to evaluate, the best Taylor series method known is better than the best nonlinear Runge-Kutta method known. However, if the cost of evaluating the k^{th} derivative of v increases faster than O(ln k) as $k \uparrow \infty$, then it is easy to show that the opposite will be true.

Section 7

Numerical Results

In the previous sections, we computed (for several classes of methods) that order which minimized the work required to attain a given error criterion. Here, we consider actual numerical results of optimal order and minimal cost for various test problems and classes of methods. The optimal order for a given error criterion was determined by finding, for each method implemented, the coarsest mesh that allowed the error criterion to be satisfied; the resulting complexities were then compared to determine the optimal order. The error measure used was the "endpoint error," i.e., the ∞-norm (see e.g., Stewart [73], pg. 164) of the difference between the true and computed solutions, evaluated at the endpoint of the interval of interest (the unit interval I). All testing was carried out on the Carnegie-Mellon University Computer Science Department's PDP-10 in ALGOL and FORTRAN, using double precision.

The first problems considered were of the form

$$\dot{x}(t) = \lambda x(t) \qquad x(0) = 1 \tag{7.1}$$

on the unit interval I. Although this problem is easy to handle analytically, any general problem of the form (2.1) may be locally approximated by a linear system of ordinary differential equations (see e.g., Hindmarsh [74], pp. 17-18). If the coefficient matrix of this linear system is diagonalizable, an uncoupled set of scalar equations of the form (7.1) will result.

These problems were solved via Taylor series methods; the optimal order is given in Table 7.1 for the choices of λ indicated. Here the optimal order was taken to

be that order which minimizes the number of evaluations of the right-hand side of (7.1) required to attain the desired error criterion. As expected, the number of evaluations required increases as the error criterion ε decreases. Moreover, the optimal order also increases monotonically as ε decreases, just as the theory predicts.

We next turn to the solution of the test problem

$$\dot{x}(t) = \cos^2 x(t) \qquad x(0) = 0 \ . \tag{7.2}$$

For this problem, we searched for the optimum "unmodified" Brent-Runge-Kutta method. For this problem, the optimal order was taken to be that for which the actual CPU time (in milliseconds) required to solve the problem to within a given ε was minimized. Since there is a certain amount of randomness in such a measure, the mean time for ten runs was analyzed. Not surprisingly, it turned out that the order which minimized the CPU time also minimized the number of evaluations of the right-hand side of (7.2). Since the $(n + 2)^{th}$-order method requires the zeros of the Jacobi polynomial $G_n(2, 2, \cdot)$, and the best set of values available only contained the zeros for $1 \le n \le 8$ (Table 25.8 of Abramowitz and Stegun [64]), only the methods of order not exceeding ten were implemented.

The results for problem (7.2) are given in Table 7.2. Here, the optimal order p^*, the optimal number of mesh points n^*, the minimal number of evaluations C_e^*, and the minimal mean CPU time C_t^* are given. Note that these all behave as predicted. In addition, we computed the ratio of the mean CPU time for a fourth-order method $C_t^*(4, \cdot)$ to the minimal mean runtime. As the theory predicts, this ratio appears to be increasing without bound as ε tends to zero. (The same behavior was found for the ratio $C_e(4, \cdot) / C_e^*$, where $C_e(4, \cdot)$ is the number of evaluations required by a fourth-order method.)

Finally, we looked at the "hard" problem

$$(7.3)\qquad \begin{aligned} \dot{x}_i(t) &= \Sigma_{j=1}^{2} \alpha_{ij}(t)\, x_i(t)\, x_j(t) \qquad (1 \le i \le 2) \\ \alpha_{ij}(t) &= \gamma_{ij} \int_{-\infty}^{+\infty} \exp(-\nu_{ij}(t-\tau)^2)\, \tau^{-1} \sin\tau \, d\tau \qquad (1 \le i, j \le 2) \end{aligned}$$

(where "exp" denotes the exponential function), with initial conditions

$$x_1(0) = x_2(0) = 1 \ .$$

The γ_{ij} were all taken to be one, while the ν_{ij} were taken to be

$$\nu_{11} = 1 \ , \quad \nu_{12} = \nu_{22} = 10^{-6} \ , \quad \nu_{21} = 10^{-3} \ .$$

(This system of differential equations is similar to the system governing a two-species gas chemical reaction; see e.g., Finlayson [71].)

Since the system (7.3) is nonscalar and nonautonomous, the Brent-Runge-Kutta methods are not appropriate. Since the derivatives of $x_i(t)$ are not readily available, the Taylor series methods are not particularly easy to apply. Thus we used linear Runge-Kutta methods for the solution of (7.3). The particular methods RKp of order p ($1 \le p \le 8$) used were as follows.

RK1 ... Euler's method
RK2 ... Ralston [66] (5.6-40) "modified Euler"
RK3 ... Ralston [66] (5.6-45)
RK4 ... Ralston [66] (5.6-48) "classical method"
RK5 ... Cassity [66]
RK6 ... Butcher [64b] (first method on page 192)
RK7 .. Shanks [66]
RK8 ... Cooper and Verner [72]

The methods of order less than eight have the optimal number of stages per step, while the method of Cooper and Verner has the minimum number of stages of all eighth-order methods known (see Section 5.2).

Most of the work involved in solving (7.3) was in evaluating $\alpha_{ij}(t)$. An obvious change of variable reduces this to a Gauss-Hermite quadrature; a twenty-point quadrature (Table 25.10 of Abramowitz and Stegun [64]) was used for maximal accuracy. The Chebyshev rational function approximation given on page 356 of

Fröberg [69] was used to compute (sin r) / r for $|r| \le 1$; the system double-precision sine routine was used for $|r| > 1$.

Since so much of the time required to solve (7.3) was spent in evaluating $\alpha_{ij}(t)$, the measure of cost was the number of evaluations of the set $\{\alpha_{ij}(t) : 1 \le i, j \le 2\}$; that is, we measured the number of evaluations of the (vector) right-hand side of (7.3). (Moreover, the amount of computer time required to search for the optimum was so great as to preclude running the problem a large number of times and averaging the results, as was done in the previous example.) Results are given in Table 7.3, where p^*, n^*, and C_e^* (defined as for (7.2)) are given as a function of the error criterion. The table stops at $\varepsilon = 10^{-5}$, since the eighth-order method (i.e., the highest-order method implemented for testing) was reached at that level. Again, note that the theoretical results predicted are confirmed in this difficult example.

So, our three numerical examples yield data which agree with the theoretical result that the optimal order $p^*(\alpha)$ increases with $\alpha = \ln \varepsilon^{-1}$. Moreover, in Sections 5 and 6, we saw that $p^*(\alpha) = O(\alpha)$ as $\alpha \uparrow \infty$; i.e., the optimal order increases linearly with α. The data in Tables 7.1-7.3 support this result.

Further testing still remains to be done. In these examples, we picked problems that were well-suited to one particular type of method (e.g., it was easy to get the derivative information required by Taylor series methods for (7.1)). Future testing should look at problems that are "neutral" in the sense that the informations required for the various classes of methods are equally hard to obtain. This would allow the comparison of various classes of methods. In addition, we point out that "fast" methods of polynomial manipulation were not used (due to the additional programming involved in designing such a package); perhaps such a package should be implemented for future testing of the Taylor series and nonlinear Runge-Kutta methods.

TABLE 7.1

Taylor Series Methods for Test Problem

$\dot{x}(t) = \lambda x(t) \quad x(0) = 1$

$-\log_{10}\varepsilon$	$\lambda = -e$	$\lambda = -1$	$\lambda = -1/e$	$\lambda = 1/e$	$\lambda = 1$	$\lambda = e$
1	2	3	1	1	3	8
2	9	4	2	2	4	9
3	11	6	3	3	6	11
4	12	7	4	4	7	12
5	14	8	5	5	8	14
6	15	9	6	6	9	15
7	16	10	7	7	10	16
8	17	11	8	8	11	18
9	19	12	9	9	12	19

Notes:

1. In all cases except $\lambda = -e$, $\varepsilon = 10^{-1}$, the optimal mesh-size was $h = 1.0$; for this exceptional case, it was $h = 0.5$.

2. Entry in table is the optimal order for the given λ and ε. This equals the minimal number of function evaluations required to solve the problem on the entire unit interval, except for the exceptional case noted above, where four was the minimal number of evaluations.

TABLE 7.2

Brent-Runge-Kutta Methods for Test Problem

$\dot{x}(t) = \cos^2 x(t) \quad x(0) = 0$

$-\log_{10}\varepsilon$	p^*	n^*	C_e^*	C_t^*	$C_t(4,\cdot)/C_t^*$
1	1	2	2	2.789	3.93
2	2	2	4	7.824	3.28
3	4	2	6	23.144	1.88
4	5	2	8	32.401	1.38
5	6	2	10	46.837	1.87
6	7	2	12	60.978	2.15
7	8	2	14	75.613	3.18
8	9	2	16	92.852	4.58
9	10	2	18	108.632	6.85

TABLE 7.3

Linear Runge-Kutta Methods for Test Problem

$$\dot{x}(t) = \sum_{j=1}^{2} \alpha_{ij}(t)\, x_i(t)\, x_j(t) \qquad x_i(0) = 1 \qquad (1 \le i \le 2)$$

$$\alpha_{ij}(t) = \gamma_{ij} \int_{-\infty}^{+\infty} \exp(-\gamma_{ij}\,(t - \tau)^2)\, \tau^{-1} \sin \tau \; d\tau \qquad (1 \le i, j \le 2)$$

$-\log_{10}\varepsilon$	p^*	n^*	C_e^*
1	3	8	24
2	4	10	40
3	4	15	60
4	7	9	81
5	8	9	99

Section 8

Summary and Conclusions

In this thesis, we have constructed a methodology for studying the computational complexity of one-step methods for the numerical solution of ordinary differential equation initial-value problems. We developed lower and upper bounds on the complexity of a given method, and showed how to pick that method within a basic sequence which minimizes complexity. Under very general hypotheses (which were later verified for a number of commonly-used classes of methods), we saw that the optimal order increases as the error criterion ε decreases, tending to infinity as ε tends to zero; this is in contrast to the situation in iterative complexity (Traub and Woźniakowski [76]). Moreover, in many of the specific classes of methods studied, we saw that the optimal step-size does not tend to zero with ε, a result indicating that it is important to not assume that h tends to zero. These results of optimal order and step-size were then used to find bounds on the complexity of solving the equation to within a given ε, using a given class of methods; using these bounds, we were able to compare the "goodness" of several such classes.

We now turn to some issues that have been raised by this study. Probably the most important point is that we have found evidence that high-order methods may be of practical (i.e., computational), as well as theoretical, interest. However, we need to learn much more about them, the crucial point not being that of getting maximal order for a given information set, but that of getting minimal complexity for a method of given order. For example, the optimally-ordered class Φ_{MBRK} has greater complexity

than the class Φ_{BRK}; indeed, the differences in combinatory cost far outweigh any advantages the former has over the latter, as was pointed out in Section 6. (Even if we were to take the drastic step of ignoring combinatory cost, we would still be faced with the fact that Φ_{BRK} is known to be order-convergent, while no such result is known for Φ_{MBRK}.) A similar situation arises when we compare the classes Φ_{CRK} and Φ_{CVRK} of linear Runge-Kutta methods (Section 5.2).

Finally, we consider some open questions.

(1.) In a number of instances, we have only been able to show "trivial" lower bounds, i.e., lower bounds which are linear in the amount of information needed. However, the best algorithms known have complexity which grows faster than linearly in the size of the information set. How may we narrow this gap? (Note that this is an issue that touches almost all areas of complexity theory.)

(2.) What are the possibilities of extending this analysis to include "polyalgorithms" for the solution of initial-value problems? It is often wise to vary the step-size from step to step; furthermore, many existing programs allow the order to vary. It would be useful to have a complexity theory that includes these methods.

(3.) We assumed throughout this thesis that infinite-precision real arithmetic was available. It would be of great interest to study the complexity of initial-value problems using variable-precision arithmetic, a far more realistic model.

(4.) What is the complexity of using multistep methods to solve initial-value problems? We have some preliminary results for the class of Adams-Bashforth predictor/Adams-Moulton corrector methods, using Φ_{BRK} to find the necessary starting values; this work will be reported in a future paper. (This class is order-convergent, and a result similar to Theorem 3.3 holds; it also appears that the choice of starting

method is of great importance in the complexity analysis.) Besides the multistep methods, many oth. classes of methods remain to be analyzed, such as extrapolation methods, spline methods, multistep Runge-Kutta methods, and special methods for "stiff" equations. (Of course this list is by no means exhaustive; see Gear [71] or Hindmarsh [. 4] for further discussion.)

(5.) The error equation (3.1), (3.2) holds for non-iterative methods for the solution of a number of other problems such as boundary-value problems for ordinary and partial differential equations. Hence, we suspect that a great deal of the analysis in Sections 2, 3, and 4 may go through unchanged. A long-term goal is the study of such problems from a complexity viewpoint.

Appendix A

Error Bounds for a Basic Sequence of Cooper-Runge-Kutta Methods

In this Appendix, we describe a subclass of a class of linear Runge-Kutta ("LRK") methods due to Cooper [69]. We shall first prove the following

Theorem A.1: There is a basic sequence $\Phi_{CRK'}$ of LRK methods such that

(1.) Each $\varphi_p \in \Phi_{CRK'}$ requires

$$s(p) := (p^2 - p + 2) / 2$$

evaluations of v per step.

(2.) There exists an $M_U > 0$ such that

(A.1) $$\sigma_G(\varphi_p, h) \leq (M_U \ln (p+e)\, h)^p$$

for $h \leq h_p = O((\ln p)^{-1})$.

We use the notation of Cooper and Verner [72]. Let $p \in \mathbb{Z}^{++}$ be given; define $\rho: \mathbb{Z}^+ \cap [0, p] \to \mathbb{Z}^+$ by

(A.2) $$\rho(j) := \begin{cases} \Sigma_{k=0}^{j} k \equiv j(j+1) / 2 & \text{if } j \neq p \\ s & \text{if } j = p \ , \end{cases}$$

where we write "s" for "s(p)" as defined above. Next, a set $\{\xi_0 , \dots , \xi_s\}$ of integers is defined by picking $\xi_0 := p$, and setting ξ_i $(i \neq 0)$ to be the unique integer in $[1, p]$ satisfying

(A.3) $$\rho(\xi_i - 1) < i \leq \rho(\xi_i) \ .$$

We now pick $u_0 , \dots , u_s \in I$ satisfying

(A.4) $$u_0 = 0 \ , \ u_s = 1 \ , \ u_i \neq 0 \text{ if } i \neq 0$$

and

(A.5) $\qquad (\xi_i = \xi_j \text{ and } i \neq j) \text{ implies } u_i \neq u_j$.

Finally, we pick a matrix of coefficients $A := \{\lambda_{ij}: 0 \le j \le i-1, 1 \le i \le s\}$ such that

(A.6) $\qquad \lambda_{ij} = 0 \quad \text{if } \xi_i < \xi_j - 1 \qquad (1 \le i, j \le s)$

and

(A.7) $\qquad \Sigma_{j=0}^{i-1} \lambda_{ij} u_j^{\tau} = (\tau+1)^{-1} u_i^{\tau+1} \qquad (0 \le \tau \le \xi_i - 1, 1 \le i \le s)$.

Cooper and Verner [72] point out that these conditions may always be fulfilled; the resulting A defines a p^{th}-order LRK method with s stages.

We are interested in a choice of $u_0, \ldots, u_s$ which will give a small error coefficient. To this end, we will choose

(A.8) $\qquad \{u_j: \xi_j = n\} = \{(1 + x_{kn}) / 2 : 1 \le k \le n\} \qquad (1 \le n \le p - 1)$,

where $x_{1n}, \ldots, x_{nn}$ are the zeros of the Jacobi polynomial $P_n := P_n^{(1,1)}$ (see Szegö [59]). Since these zeros are distinct and lie in $[-1, 1]$, conditions (A.4) and (A.5) may be satisfied.

Now we are able to exhibit a solution to the i^{th} system in (A.7). First, note that the equation for $\tau = 0$ may be separated from the others, since $u_0 = 0$. Setting

$$n := \xi_i - 1 ,$$

we see that

(A.9) $\qquad \lambda_{i0} = u_i - \Sigma_{j=1}^{i-1} \lambda_{ij} = u_i - \Sigma \{ \lambda_{ij}: j < i \text{ and } \xi_j \ge n \}$,

the last by (A.6). We wish to determine the nonzero λ_{ij}, i.e., those λ_{ij} for which $\xi_j \ge n$ and $j < i$. So setting

$$\lambda_{ij} = 0 \quad \text{unless} \quad j \in \{j_1, \ldots, j_n\} ,$$

we see that the remaining λ_{ij} are the solution of the system

$$\Sigma_{k=1}^{n} u_{j_k}^{\tau} \lambda_{ij_k} = (\tau+1)^{-1} u_i^{\tau+1} \qquad (1 \le \tau \le n) .$$

Thus the λ_{ij_k} are the weights for an interpolatory quadrature formula on $[0, u_i]$ with

abscissae $u_{j_1}, \dots, u_{j_n}$. From the usual expression for such weights and (A.6), we see that

$$\lambda_{ij_k} = \mu_{ikn} := [2P_n'(\cos \vartheta_{kn})]^{-1} \int_{\vartheta_{i,n+1}}^{\pi} [P_n(\cos \vartheta) / (\cos \vartheta - \cos \vartheta_{kn})] \sin \vartheta \, d\vartheta \ ,$$

where $x_{kn} = \cos \vartheta_{kn}$ $(1 \le k \le n)$.

Lemma A.1: $\mu_{ikn} = O(n^{-1} \ln n)$ as $n \uparrow \infty$.

Proof: Since the zeros of P_n are symmetric about the origin, we may assume that $0 < \vartheta_{kn} \le \pi/2$. Using (8.9.2) of Szegö [59], we then find

$$\mu_{ikn} = O(k^{5/2} n^{-3}) \int_{\vartheta_{i,n+1}}^{\pi} [P_n(\cos \vartheta) / (\cos \vartheta - \cos \vartheta_{kn})] \sin \vartheta \, d\vartheta \ .$$

Case 1: $\vartheta_{1,n+1} \le \vartheta_{i,n+1} \le \vartheta_{k,n+1}/2$. We consider the integral over $[\vartheta_{1n}/2, \vartheta_{i,n+1}]$, since Theorem 15.4 of Szegö [59] proves that

$$O(k^{5/2} n^{-3}) [| \int_0^{\pi} | + | \int_0^{\vartheta_{1n}/2} |] = O(n^{-1}) \ ;$$

here the integrand is the same as in the preceding integral. But the proof of (15.4.12) in Szegö [59] extends almost immediately to a proof that the remaining integral is $O(k^{-2} n)$, since (15.4.12) is proved by order-of-magnitude estimates. Thus $\mu_{ikn} = O(n^{-1}) = O(n^{-1} \ln n)$ for Case 1.

Case 2: $\vartheta_{k,n+1}/2 \le \vartheta_{i,n+1} \le 3\vartheta_{k,n+1}/2$. We consider the integral over $[\vartheta_{kn}/2, \vartheta_{i,n+1}]$, since Szegö [59] shows that

$$O(k^{5/2} n^{-3}) \, | \int_{\vartheta_{kn}/2}^{\pi} | = O(n^{-1}) \ .$$

As in (15.4.13) of Szegö [59], we have

$$\int_{\vartheta_{kn}/2}^{\vartheta_{i,n+1}} = O(nk^{-3/2}) \, I_1 + I_2 \ .$$

Here

$$I_1 := \int_{\vartheta_{kn}/2}^{\vartheta_{i,n+1}} D(\vartheta) \sin \vartheta \, d\vartheta \ ,$$

with

$$D(\vartheta) := [\cos (N\vartheta + \gamma) - \cos (N\vartheta_{kn} + \gamma)] / [\cos \vartheta - \cos \vartheta_{kn}] \ ,$$

where $N := n + 3/2$ and $\gamma := -3\pi/4$, and

$$I_2 := \int_{\vartheta_{kn}/2}^{\vartheta_{i,n+1}} R_n(\vartheta,\vartheta_{kn}) \sin\vartheta \, d\vartheta = O(nk^{-3/2}),$$

with R_n the remainder term in (8.8.2) of Szegő [59]. Unfortunately, the proof that (15.4.14) of Szegő [59] is bounded does not extend to a proof that I_1 is bounded, since the proof of the former requires that the interval of integration be symmetric about ϑ_{kn}. However, it is straightforward to verify that

$$I_1 = O(1) \int_0^{\pi/4} |\sin N\vartheta / \vartheta| \, d\vartheta = O(\ln n).$$

Thus $\mu_{ikn} = O(n^{-2}k \ln n) = O(n^{-1} \ln n)$ for Case 2.

Case 3: $3\vartheta_{k,n+1} \le \vartheta_{i,n+1} \le 3\pi/4$. We consider the integral over $[3\vartheta_{kn}/2, \vartheta_{i,n+1}]$, since Szegő [59] proves that

$$O(k^{5/2}n^{-3}) \left| \int_{3\vartheta_{kn}/2}^{\pi} \right| = O(n^{-1}).$$

But the proof of (15.4.19) in Szegő [59] extends to prove that the remaining integral is $O(k^{-5/2}n)$ (as in Case 1). Thus $\mu_{ikn} = O(n^{-1}) = O(n^{-1} \ln n)$ for Case 3.

Case 4: $3\pi/4 \le \vartheta_{i,n+1} \le \vartheta_{n+1,n+1}$. We consider the integral over $[3\pi/4, \vartheta_{i,n+1}]$, since Szegő [59] shows that

$$O(k^{5/2}n^{-3}) \left| \int_{3\pi/4}^{\pi} \right| = O(n^{-1}).$$

As in Cases 1 and 3, the proof of the above may be extended to prove a similar bound on the integral of interest. Thus $\mu_{ikn} = O(n^{-1}) = O(n^{-1} \ln n)$ in Case 4, completing the proof of the Lemma ∎

Thus (A.9) and Lemma A.1 show the existence of a $\lambda > 0$ such that

$$\Sigma_{j=0}^{i-1} |\lambda_{ij}| \le \lambda \ln(\xi_i + e); \tag{A.10}$$

here λ is independent of p. Moreover, the result for the case $i = s$ may be sharpened. We see that $\lambda_{sj} \ge 0$, since the u_j for the s^{th} system in (A.7) are the abscissae for Lobatto quadrature. Thus

$$\Sigma_{j=0}^{s-1} |\lambda_{sj}| = \Sigma_{j=0}^{s-1} \lambda_{sj} = 1, \tag{A.11}$$

the consistency condition in the last equality being a consequence of (A.7) with $r = 0$.

Proof of Theorem A.1: As in Cooper and Verner [72], we define

$$\varepsilon_i := \dot{x}(u_i h) - k_i$$

and

$$\delta_i := \int_0^{u_i} \dot{x}(uh)\, du - \Sigma_{j=0}^{i-1} \lambda_{ij}\, \dot{x}(u_j h)$$

for $0 \le i \le s$; note that $\delta_0 = \varepsilon_0 = 0$. Let $z(h)$ be the computed approximation to $x(h)$; then

$$\begin{aligned} h^{-1} \|x(h) - z(h)\| &= \| h^{-1} [x(h) - x(0)] - \Sigma_{i=0}^{s-1} \lambda_{si}\, k_i \| \\ &\le \|\delta_s\| + \|\Sigma_{i=0}^{s-1} \lambda_{si}\, \varepsilon_i\| \\ &\le \|\delta_s\| + \max_{\xi_i = p-1} \|\varepsilon_i\| , \end{aligned} \tag{A.12}$$

the last by (A.6) and (A.11). By the analyticity of x, there is an $A_1 > 0$ such that

$$\beta_i := h^{-1} \| x(u_i h) - \Sigma_{\tau=0}^{\xi_i} (u_i h)^\tau x^{(\tau)}(0) / \tau! \| \le (A_1 h)^{\xi_i}$$

and

$$\gamma_{ij} := \| \dot{x}(u_j h) - \Sigma_{\tau=0}^{\xi_i - 1} (u_j h)^\tau \dot{x}^{(\tau)}(0) / \tau! \| \le (A_1 h)^{\xi_i} ,$$

so that the definition of δ_i gives

$$\begin{aligned} \|\delta_i\| &\le \beta_i + \Sigma_{j=0}^{i-1} |\lambda_{ij}|\, \gamma_{ij} \\ &\le (A_1 h)^{\xi_i} + \Sigma_{j=0}^{i-1} |\lambda_{ij}|\, (A_1 h)^{\xi_i} \\ &\le (A_2 h)^{\xi_i} \end{aligned} \tag{A.13}$$

for a suitable $A_2 > 0$. Thus (A.12) becomes

$$h^{-1} \|x(h) - z(h)\| \le (A_2 h)^p + \max_{\xi_i = p-1} \|\varepsilon_i\| . \tag{A.14}$$

We now use Lemma 1.1 of Cooper and Verner [72] and (A.6) to find that if L is a Lipschitz constant for v, then there exists $A_3 > 0$ such that

$$\begin{aligned} \|\varepsilon_i\| &\le hL \|\delta_i\| + hL \Sigma_{j=0}^{i-1} |\lambda_{ij}| \max_j \|\varepsilon_j\| \\ &\le (A_3 h)^{\xi_i + 1} + (A_3 h) \ln (\xi_i + e) \max_j \|\varepsilon_j\| , \end{aligned}$$

the last by (A.10) and (A.13); here, the maximum is taken over all $j < i$ such that $\xi_j \ge \xi_i - 1$. A straightforward induction shows that if $(1 + \ln 2) A_3 h < 1$, then

$$\|s_i\| \le (A_4 \ln(\xi_i + e)\, h)^{\xi_i+1}$$

for a suitable $A_4 > 0$. Combining this with (A.14), we find

$$h^{-1}\,\|x(h) - z(h)\| \le (A_5 \ln(p+e)\, h)^p, \tag{A.15}$$

the desired bound for the local error for a single unit step.

To extend (A.15) to a global error result, we must look at the Lipschitz constants for the increment functions. Let L be a bound on $\|\nabla v\|$, and write "$\nabla\varphi_p(y,h)$" to indicate gradient with respect to the vector variable y. Now

$$\|\nabla\varphi_p(y,h)\| \le \Sigma_{i=0}^{s-1} |\lambda_{si}| \max_{0\le i\le s-1} \|\nabla k_i(y,h)\|$$
$$= \max_{0\le i\le s-1} \|\nabla k_i(y,h)\|,$$

where we write "$k_i(y,h)$" to indicate the dependence of k_i upon y and h. By the definition of $k_i(y,h)$, we find

$$\nabla k_i(y,h) = \nabla v(u)\,[1_{N\times N} + h\,\Sigma_{j=0}^{i-1} \lambda_{ij}\,\nabla k_j(y,h)],$$

where $u := y + h\,\Sigma_{j=0}^{i-1} \lambda_{ij}\, k_j(y,h)$ and $1_{N\times N}$ is an N×N identity matrix. Taking norms in the above gives the result

$$\gamma_i \le L\lambda + hL\lambda\,[\ln(\xi_i+e) \max\{\gamma_j:\ j < i \text{ and } \xi_j \ge \xi_i - 1\}],$$

where $\gamma_i := \|\nabla k_i(y,h)\|$. Writing λ_p for the Lipschitz constant for φ_p, it is easy to see that (A.16) and the above inequality imply

$$\lambda_p \le \Sigma_{j=0}^{p-1} (hL\lambda)^j \Pi_{k=1}^{j-2} \ln(p+e-k),$$

which is bounded for all p, provided that $h \le h_p < (L\lambda \ln(p+e))^{-1}$. Thus (A.1) follows from this result, (A.15), and Theorem 3.3 of Henrici [62]. ∎

The value for s(p) indicated in Theorem A.1 may be improved somewhat by noting that since we are using a Lobatto quadrature, higher order may be expected with fewer steps. Indeed, if we use the strategy outlined in the comments following Theorem 4 of Cooper and Verner [72], we have

Theorem A.2: There exists a basic sequence Φ_{CRK} of LRK methods such that (A.1) holds and φ_p requires

$$s(p) := \lfloor (p^2 - 2p + 4) / 2 \rfloor$$

evaluations of v per step. ∎

Appendix B

Order-Convergence of a Basic Sequence of Brent-Runge-Kutta Methods

In this Appendix, we describe a subclass of a class of iterative methods for the solution of scalar nonlinear equations. This subclass will then be used to generate an order-convergent basic sequence Φ_{BRK} of nonlinear Runge-Kutta methods.

Lemma B.1: Let $F: D \subset \mathbb{R} \to \mathbb{R}$ have a simple zero ξ, and suppose that F is analytic at ξ. Pick $k, m \in \mathbb{Z}^{++}$ with $m + 1 \geq k$. Then there is a sequence $\Psi_{km} := \{\psi_{kmn} : n \in \mathbb{Z}^{++}\}$ of stationary multipoint methods without memory such that the following hold:

(1.) The method ψ_{kmn} uses the information

$$\mathfrak{I}_{kmn} := \{F(x_0), \ldots, F^{(m)}(x_0), F^{(k)}(y_1), \ldots, F^{(k)}(y_n)\}$$

(the points $y_1, \ldots, y_n$ being suitably chosen) to compute a new approximation x_1 to ξ from a given approximation x_0 by setting

$$x_1 := \psi_{kmn}(x_0) .$$

(2.) There exists a $B > 0$ and an $h_0 > 0$ such that if $|x_0 - \xi| \leq h_0$, then

$$|x_1 - \xi| \leq (B\,|x_0 - \xi|)^p \quad \text{for all } n \in \mathbb{Z}^{++},$$

where

$$p := \min (m + 2n + 1, 2m + n + 1) . \tag{B.1}$$

Before proving the Lemma, we describe how the method ψ_{kmn} computes an improved approximation x_1 from the old approximation x_0.

Algorithm for computing $x_1 := \psi_{kmn}(x_0)$.

(1.) Let $\delta := |F(x_0)/F'(x_0)|$.

(2.) Let z_1 be an approximate zero of

$$p_1(x) := \Sigma_{i=0}^{m} (x - x_0)^i F^{(i)}(x_0) / i!$$

satisfying

(B.2) $$z_1 = x_0 + O(\delta) \quad \text{and} \quad |p_1(z_1)| \le (A_1 \delta)^{m+1},$$

where A_1 is independent of n.

(3.) Let

$$y_i := x_0 + \alpha_{in} (z_1 - x_0) \quad (1 \le i \le n),$$

where

$$\alpha_{in} := (1 + x_{i+n}) / 2$$

and $x_{1n} > \ldots > x_{nn}$ are the zeros of the Jacobi polynomial

$$P_n(x) := p_n^{(k-1,\, m+1-k)}(x)$$

(see Szegő [59]).

(4.) Let p_{n+1} be the polynomial of degree at most m + n that interpolates the information $\mathfrak{N}_{kmn}$, and let x_1 be an approximate zero of p_{n+1} satisfying

(B.3) $$x_1 = x_0 + O(\delta) \quad \text{and} \quad |p_{n+1}(x_1)| \le (A_2 \delta)^{\rho},$$

where A_2 is independent of n and ρ is given by (B.1).

Here we use the notation of Brent [74]. Clearly, $\psi_{kmn} \in C'(k, m, n)$, the only difference being that conditions (B.2) and (B.3) replace (2.2) and (2.4) of Brent [74]. It is easy to see that (B.2) and (B.3) may be realized by using $\lceil \log_2(m+1) \rceil - 1$ and $\lceil \log_2(\rho/(m+1)) \rceil$ iterations of Newton's method, with the respective starting approximations of $x_0 - F(x_0) / F'(x_0)$ and z_1.

Proof of Lemma B.1: Let x_1' be the exact zero of p_{n+1} near x_0. We then find that there is a ξ between x_1' and z_1 such that

$$(B.4)\qquad |F(x_1')| \le |p_{n+1}(z_1) - F(z_1)| + |p_{n+1}'(\xi) - F'(\xi)|\,|x_1' - z_1| .$$

Using (B.3), the analyticity of F, and standard techniques of interpolation theory (Traub [64]), it is easy to show that (2.9) and (2.10) of Brent [74] may be rewritten as

$$(B.5)\qquad \begin{aligned} |p_{n+1}(x) - F(x)| &\le (A_3\,\delta)^{m+n+1} \quad \text{and} \\ |p_{n+1}'(x) - F'(x)| &\le (A_4\,\delta)^{m+n} \end{aligned}$$

for $|x - x_0| \le 4\delta$. (Here all constants A_r will be independent of n.) Similarly, we find that

$$|x_1' - \xi| \le (A_5\,\delta)^{m+n} \quad \text{and} \quad |z_1 - \xi| \le (A_6\,\delta)^{m+1} ,$$

so that the triangle inequality gives

$$(B.6)\qquad |x_1' - z_1| \le (A_7\,\delta)^{m+1} .$$

Using (B.4), (B.5), and (B.6), we see that

$$(B.7)\qquad \begin{aligned} |f(x_1')| &\le |p_{n+1}(z_1) - F(z_1)| + (A_8\,\delta)^{2m+n+1} \\ &\le |p_{n+1}(z_1) - F_1(z_1)| + |F_2(z_1)| + (A_8\,\delta)^{2m+n+1} , \end{aligned}$$

where

$$F_1(x) := \sum_{i=0}^{m+2n} (x - x_0)^i F^{(i)}(x_0) / i! \quad \text{and} \quad F_2(x) := F(x) - F_1(x) .$$

Clearly $|F_2(x)| \le (A_9\,\delta)^{m+2n+1}$, so that (B.7) becomes

$$(B.8)\qquad |F(x_1')| \le |p_{n+1}(z_1) - F_1(z_1)| + (A_{10}\,\delta)^{\theta} .$$

As in Brent [74], we now write

$$p_{n+1}(x) = r_1(x) + r_2(x) ,$$

where r_i (i = 1, 2) is the polynomial of degree at most m + n satisfying

$$r_i^{(j)}(x_0) = F_i^{(j)}(x_0) \quad (0 \le j \le m)$$

and

$$r_i^{(k)}(y_j) = F_i^{(k)}(y_j) \quad (1 \le j \le n) .$$

If we let

$$P(x) := r_1(x + x_0) - F_1(x + x_0) ,$$

and write $\epsilon := z_1 - x_0$ (in this Appendix only), we find that

$$P^{(i)}(0) = 0 \quad (0 \le i \le m) \quad \text{and} \quad P^{(k)}(\alpha_{in}\,\epsilon) = 0 \quad (1 \le i \le n) .$$

We may easily alter the proof of Lemma 4.3 in Brent [74] to show that

$$r_1(z_1) - F_1(z_1) = P(\epsilon) = 0 .$$

Thus (B.8) becomes

$$|F(x_1')| \le |r_2(z_1)| + (A_{10}\,\delta)^{\rho} . \tag{B.9}$$

To bound the remaining term, let us write

$$r_2(x) = \Sigma_{j=1}^{m} a_{j+m} (x - x_0)^{j+m} ,$$

recalling that r_2 has a zero of multiplicity m at x_0. Using the notation of Stewart [73], we see that the nonzero coefficients of r_2 are given by the solution of the linear system

$$Wy = c ,$$

where

$$\omega_{ij} := \alpha_{in}^{j-1} \quad (1 \le i, j \le n) ,$$

$$\xi_j := a_{j+m}\,\epsilon^{j+m}\,(j+m)! / (j+m-k)! \quad (1 \le j \le n) , \text{ and}$$

$$\gamma_i := \epsilon^k F_2^{(k)}(y_i) / \alpha_{in}^{m-k+1} \quad (1 \le i \le n) .$$

Since W^T is a Vandermonde matrix, we find that the entries of $U = W^{-1}$ are given by

$$\psi_{ij} = \alpha_{jn}\,(-1)^{n-i}\,\sigma_{n-i,n-1,j} / \Pi_{r \ne j} (\alpha_{jn} - \alpha_{rn}) ,$$

where

$$\sigma_{\mu,n-1,j} := \Sigma\ \alpha_{p_1,n} \cdots \alpha_{p_\mu,n} ,$$

the sum being taken over all multi-indices $p_1 \ldots p_\mu$ not including j (Gregory and Karney [69]). Since there are fewer than 2^n summands, each of which lies in [0, 1], we see that $\sigma_{\mu,n-1,j} \le 2^n$, implying that

$$|\psi_{ij}| \le 2^n \alpha_{jn} / \Pi_{r \ne j} (\alpha_{jn} - \alpha_{rn}) .$$

So we have

$$(8.10)\qquad \begin{aligned} |\eta_i| &= |\Sigma_{j=1}^{n} v_{ij}\gamma_j| \\ &\le n\,2^n \max_{1\le j\le n} |\delta^k F_2^{(k)}(y_j) / [\alpha_{jn}^{m-k} G_n'(\alpha_{jn})]| , \end{aligned}$$

where

$$G_n(x) := G_n(m+1, m+2-k, x) = \Pi_{r=1}^{n} (x - \alpha_{rn})$$

(see Abramowitz and Stegun [64]).

Now it is clear that

$$\max_{1\le j\le n} 1 / \alpha_{jn}^{m-k} = 1 / \alpha_{nn}^{m-k} .$$

By Theorem 8.9.1 of Szegö [59], we may show that

$$\alpha_{nn} \ge A_{11}\, n^{-2} ;$$

using this result and (22.5.2) of Abramowitz and Stegun [64], we find that

$$(8.11)\qquad \begin{aligned} &\max_{1\le j\le n} [\alpha_{jn}^{m-k} G_n'(\alpha_{jn})]^{-1} \\ &\le A_{12}\, n^{2(m-k)} \binom{m+2n+1}{m} \max_{1\le j\le n} |P_n'(x_{jn})|^{-1} , \end{aligned}$$

By the symmetry relation (4.1.3) of Szegö [59], we may assume that $0 \le x_{jn} < 1$. Using Theorem 8.9.1 of Szegö [59], we may show that

$$|P_n'(x_{jn})|^{-1} \le (A_{13})^n ,$$

and so (8.10), (8.11), the definition of F_2, and the above imply that

$$|\eta_i| \le (A_{14}\,\delta)^{m+2n+1} ,$$

yielding the result

$$|r_2(z_1)| \le \Sigma_{j=1}^{n} a_{j+m}\, \delta^{j+m} \le n \max_{1\le i\le n} |\eta_i| \le (A_{15}\,\delta)^{m+2n+1} .$$

So (8.9) becomes

$$|F(x_1')| \le (A_{16}\,\delta)^{\rho} .$$

By Taylor's Theorem, this implies

$$|x_1' - \beta| \le (A_{17}\,\delta)^{\rho} .$$

The desired result then follows from (8.3) and from (2.5) of Brent [74]. ∎

We now describe the basic sequence Φ_{BRK}. The methods in this basic sequence are given by

$$\varphi_1(x_0, h) := v(x_0),$$

$$\varphi_2(x_0, h) := v(x_0 + h\, v(x_0)/2),$$

and for $p \geq 2$,

$$\varphi_p(x_0, h) := h^{-1}[\psi_{1,1,p-2}(x_0) - x_0],$$

with $\psi_{1,1,p-2}$ applied to the function F given by (5.1.5) and the approximation x_1 to x_1' being given by an appropriate number of iterations of Newton's method (as described above).

Theorem B.1: The basic sequence Φ_{BRK} is order-convergent with respect to the global error. Moreover, the number of stages s(p) required by $\varphi_p \in \Phi_{BRK}$ is given by

$$s(p) = \begin{cases} p & \text{if } p \leq 2 \\ p-1 & \text{if } p > 2 \end{cases}.$$

Proof: We use the notation of Lemma B.1, writing z(h) for the computed p^{th}-order approximation x_1 to x(h) and $p_{n+1}(\cdot, x_0)$ for the polynomial p_{n+1}. The result of Lemma B.1 is that

$$h^{-1}|z(h) - x(h)| \leq (8h)^p,$$

the desired result for a single unit step. To prove the global result, we must consider the Lipschitz constants for Φ_{BRK}.

We implicitly differentiate the result $p_{n+1}(x_1', x_0) \equiv 0$ to find

$$\partial_1 \varphi_p(x_0, h) = -h^{-1} Q_{n+1}(x_1', x_0) + s_p(x_0),$$

where

$$Q_{n+1}(x_1', x_0) = 1 + \partial_2 p_{n+1}(x_1', x_0) / \partial_1 p_{n+1}(x_1', x_0)$$

and

$$s_p(x_0) = h^{-1} (d/dx_0)[x_1 - x_1'].$$

It is easy to see that x_1 and x_1' are analytic functions of x_0. Since their difference tends to zero uniformly on the domain of v as $p \uparrow \infty$, it follows that

$$\lim_{p\uparrow\infty} \varepsilon_p(x_0) = 0 .$$

We claim that

$$Q_{n+1}(x_1', x_0) = O(h \ln n) \quad \text{as} \quad n \uparrow \infty ,$$

uniformly in x_0. To see this, note that we may write the interpolation polynomial p_{n+1} in terms of Jacobi polynomial P_n, finding that

$$p_{n+1}(x,x_0) = (-1)^n (h/2) \int_{-1}^{\xi(x)} P_n(t)\, dt + h\, v(x_0) \sum_{k=1}^{n} l_{kn} - h ,$$

where

$$\xi(x) := 2 (x - x_0) / [h\, v(x_0)] - 1$$

and

$$l_{kn} := [2 (1 + x_{kn})\, v(y_k)\, P_n'(x_{kn})]^{-1} \int_{-1}^{\xi(x)} (t + 1)\, P_n(t) / (t - x_{kn})\, dt .$$

Now

$$\partial_1 p_{n+1}(x_1', x_0) = (-1)^n P_n(\xi_1) / v(x_0) + (1 + \xi_1) \sum_{k=1}^{n} g(x_{kn})\, L_{kn}(\xi_1) ,$$

where

$$\xi_1 := \xi(x_1') ,$$

$$L_{kn}(x) := P_n(x) / [P_n'(x_{kn})\, (x - x_{kn})] , \text{ and}$$

$$g(t) := 1 / [(1 + t)\, v(x_0 + (1 + t)\, h\, v(x_0) / 2)] .$$

By (8.21.10) of Szegő [59], the first term in the expression for $\partial_1 p_{n+1}(x_1', x_0)$ goes to zero as $n \uparrow \infty$. A minor modification of the proof of Theorem 14.4 of Szegő [59] shows that the sum in the remaining term tends to $g(\xi(x(h)))$ as $n \uparrow \infty$. So

$$\partial_1 p_{n+1}(x_1', x_0) \sim v(x(h))^{-1} \quad \text{as} \quad n \uparrow \infty .$$

Using Lemma A.1 and techniques similar to those yielding the above estimate, we find

$$\partial_2 p_{n+1}(x_1', x_0) = O(h \ln n) - v(x(h))^{-1} \quad \text{as} \quad n \uparrow \infty .$$

This gives the estimate claimed for $Q_{n+1}(x_1', x_0)$.

So the Lipschitz constant for $\varphi_p \in \Phi_{BRK}$ grows as the logarithm of p. By Proposition 4.3, Φ_{BRK} is order-convergent. ∎

References

Abramowitz and Stegun [64]:

Abramowitz, M., and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington, D. C.: National Bureau of Standards, 1964.

Ahlfors [66]:

Ahlfors, L. V., Complex Analysis, Second Edition. New York: McGraw-Hill, 1966.

Borodin and Munro [75]:

Borodin, A. and I. Munro, The Computational Complexity of Algebraic and Numeric Problems. New York: American Elsevier, 1975.

Brent [74]:

Brent, R. P., "Efficient Methods for Finding Zeros of Functions Whose Derivatives are Easy to Evaluate," Report, Computer Science Department, Carnegie-Mellon University, 1974.

Brent [76]:

Brent, R. P., "A Class of Optimal-Order Zero Finding Methods Using Derivative Evaluations," in Analytic Computational Complexity, edited by J. F. Traub. New York: Academic Press, 1976.

Brent and Kung [76]:

Brent, R. P. and H. T. Kung, "Fast Algorithms for Manipulating Formal Power Series," Report, Computer Science Department, Carnegie-Mellon University, 1976.

Buck [65]:

Buck, R. C., Advanced Calculus, Second Edition. New York: McGraw-Hill, 1965.

Butcher [64a]:

Butcher, J. C., "Implicit Runge-Kutta Processes," Math. Comp., Vol. 18, pp. 50-64, January, 1964.

Butcher [64b]:

Butcher, J. C., "On Runge-Kutta Processes of High Order," J. Austral. Math. Soc., Vol. 4, pp. 179-194, 1964.

Butcher [75]:

Butcher, J. C., "An Order Bound for Runge-Kutta Methods," SIAM J. Num. Anal., Vol. 12, No. 3, pp. 304-315, June, 1975.

Cassity [66]:
Cassity, C. R., "Solutions of the Fifth-Order Runge-Kutta Equations," SIAM J. Num. Anal., Vol. 3, No. 4, pp. 598-606, December, 1966.

Cooper [69]:
Cooper, G. J., "Error Bounds for Some Single-Step Methods," Conf. on the Numerical Solution of Differential Equations, Lecture Notes in Mathematics 109, pp. 140-147. Berlin: Springer-Verlag, 1969.

Cooper and Verner [72]:
Cooper, G. J. and J. H. Verner, "Some Explicit Runge-Kutta Methods of High Order," SIAM J. Num. Anal., Vol. 9, No. 3, pp. 389-405, September, 1972.

Finlayson [71]:
Finlayson, B. E., "Covergence of the Galerkin Method for Nonlinear Problems Involving Chemical Reaction," SIAM J. Num. Anal., Vol. 8, No. 2, pp. 316-324, June, 1971.

Friedman [69]:
Friedman, A., Partial Differential Equations. New York: Holt, Rinehart, and Winston, 1969.

Fröberg [69]:
Fröberg, C. E., Introduction to Numerical Analysis, Second Edition. Reading: Addison-Wesley, 1969.

Gear [71]:
Gear, C. W., Numerical Initial Value Problems in Ordinary Differential Equations. Englewood Cliffs: Prentice-Hall, 1971.

Gregory and Karney [69]:
Gregory, R. T. and D. L. Karney, A Collection of Matrices for Testing Computational Algorithms. New York: Wiley-Interscience, 1969.

Gurtin [75]:
Gurtin, M. E., "Thermodynamics and Stability," Arch. Rat. Mech. Anal., Vol. 59, No. 1, pp. 63-96, 1975.

Henrici [62]:
Henrici, P., Discrete Variable Methods in Ordinary Differential Equations. New York: Wiley, 1962.

Hindmarsh [74]:
Hindmarsh, A. C., "Numerical Solution of Ordinary Differential Equations: Lecture Notes." Lawrence Livermore Laboratory Report No. UCID-16588, June, 1974.

Hull et al. [72]:
Hull, T. E., W. H. Enright, B. M. Fellen, and A. E. Sedgwick, "Comparing Numerical Methods for Ordinary Differential Equations," SIAM J. Num. Anal., Vol. 9, No. 4, pp. 603-637, December, 1972.

Kung and Traub [73]:
Kung, H. T. and J. F. Traub, "Computational Complexity of One-Point and Multipoint Iteration," in Complexity of Real Computation, edited by R. Karp. Providence: Amer. Math. Soc., 1973.

Kung and Traub [74]:
Kung, H. T. and J. F. Traub, "Optimal Order of One-Point and Multipoint Iteration." JACM, Vol. 21, No. 4, pp. 643-651, October, 1974.

Lindberg [74]:
Lindberg, B., "Optimal Stepsize Sequences and Requirements for the Local Error for Methods for Stiff Differential Equations." Technical Report No. 67, Department of Computer Science, University of Toronto, May, 1974.

Meersman [76]:
Meersman, R., "On Maximal Order of Families of Iterations for Nonlinear Equations," Doctoral Thesis, Vrije Universiteit Brussel, Brussels, 1976.

Pólya and Szegő [25]:
Pólya, G. and G. Szegő, Aufgaben und Lehrsätze der Analysis, Vol. I. Berlin: Springer-Verlag, 1925.

Rall [69]:
Rall, L. B., Computational Solution of Nonlinear Operator Equations. New York: John Wiley and Sons, Inc., 1969.

Ralston [66]:
Ralston, A., A First Course in Numerical Analysis, New York: McGraw-Hill, 1966.

Sandberg [67]:
Sandberg, I. W., "Two Theorems on the Accuracy of Numerical Solutions of Ordinary Differential Equations," Bell Sys. Tech. J., Vol. 46, pp. 1243-1266, July-August, 1967.

Shanks [66]:
Shanks, E. B., "Solutions of Differential Equations by Evaluations of Functions," Math. Comp., Vol. 20, No. 93, pp. 21-38, January, 1966.

Stetter [73]:
Stetter, H. J., Analysis of Discretization Methods for Ordinary Differential Equations. Berlin: Springer-Verlag, 1972.

Stewart [73]:
Stewart, G. W.., Introduction to Matrix Computations. New York: Academic Press, 1973.

Szegö [59]:
Szegö, G., Orthogonal Polynomials. Amer. Math. Soc. Colloquium Publications, Vol. XXIII. New York: Amer. Math. Soc., 1959.

Traub [64]:
Traub, J. F., Iterative Methods for the Solution of Equations. Englewood Cliffs: Prentice-Hall, 1964.

Traub [72]:
Traub, J. F., "Computational Complexity of Iterative Processes," SIAM J. Comput., Vol. 1, No. 2, pp. 167-179, June, 1972.

Traub and Woźniakowski [76]:
Traub, J. F. and H. Woźniakowski, "Strict Lower and Upper Bounds on Iterative Complexity," in Analytic Computational Complexity, edited by J. F. Traub. New York: Academic Press, 1976.

Woźniakowski [75]:
Woźniakowski, H, "Generalized Information and Maximal Order of Iteration for Operator Equations," SIAM J. Num. Anal., Vol. 12, No. 1, pp. 121-135, March, 1975.